

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INFÉRENCE ROBUSTE SUR LES PARAMÈTRES D'UNE RÉGRESSION
LINÉAIRE BAYÉSIENNE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR
PHILIPPE GAGNON

JANVIER 2012

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

On peut croire *a priori* que les remerciements représentent la section « facile » d'un mémoire. Ce n'est toutefois pas le cas ! En effet, on ne veut pas oublier personne sans pour autant remercier le facteur. La réflexion consacrée à la rédaction de cette section peut alors être plus intense que celle menant à une preuve mathématique...

Je tiens tout d'abord à remercier mon directeur Alain Desgagné. Il a contribué à mon développement sous plusieurs plans. Il m'a aidé à passer d'un bon étudiant à un bon chercheur. Je le remercie encore une fois pour ceci. Il m'a confié un sujet que j'ai grandement apprécié et il a su ajuster la nature du mémoire en fonction de mes sentiments. De plus, il m'a conseillé dans ma poursuite vers le doctorat et m'a épaulé dans la réalisation des tâches connexes à la recherche. Somme toute, je l'ai apprécié comme personne. Ces valeurs et son éthique implacable m'ont orienté en tant qu'étudiant, mais avant tout en tant qu'être humain. Gros merci !

Pour rester dans la lignée académique, je tiens à remercier les professeurs dont j'ai eu le plaisir de recevoir l'enseignement lors de la maîtrise. Les cours reçus m'ont servi à un moment ou à un autre dans le cadre de la recherche. Je remercie donc Mylène Bédard, Pierre Lafaye de Micheaux, Geneviève Lefebvre, Glenn Shorrock et Sorona Froda. Je remercie aussi particulièrement Manon Gauthier qui a répondu à toutes mes questions d'ordre administratif durant la maîtrise. Dans un même ordre d'idée, je remercie Gisèle Legault pour son soutien technique en laboratoire informatique.

Je veux remercier aussi ma famille proche pour m'avoir soutenu, encouragé et enduré durant les périodes de stress. Merci beaucoup Patrice, Isabelle, Émilie, Pierre, Suzanne et Bernard.

Je remercie aussi ma deuxième famille, mes amis proches, pour m'avoir eux aussi en-

couragé et enduré durant la maîtrise. Merci Marco, Jean Bertrand, Dave, Laurence et Elyse.

Merci aussi à mes *partners* pour les 5 à 7 et pour l'entraide dans les devoirs, travaux, etc. Je remercie donc Ahmed, Alexandre, Étienne, Maxime, Marie-Hélène, Mathieu, Sandra et Sadoune.

Merci à Mathieu et à Hugues pour ces conversations qui nous changeaient les idées dans le laboratoire informatique. Je crois bien avoir réussi à les pousser vers le doctorat...

Merci encore à tous ces gens qui ont de loin facilité mon passage au deuxième cycle et qui l'ont rendu si agréable !

TABLE DES MATIÈRES

LISTE DES FIGURES	vii
LISTE DES TABLEAUX	ix
RÉSUMÉ	xi
INTRODUCTION	1
CHAPITRE I	
CONTEXTE MATHÉMATIQUE	5
1.1 Présentation du modèle de régression sous sa forme usuelle	5
1.2 Le modèle de régression présenté sous l'angle d'un problème de position-échelle	6
1.3 Le modèle de régression dans un contexte d'inférence robuste	8
CHAPITRE II	
INFÉRENCE ROBUSTE	11
2.1 Famille de fonctions à variation L-exponentielle	11
2.2 Théorème de robustesse	13
CHAPITRE III	
FAMILLE DE DENSITÉS À VARIATION L-EXPONENTIELLE	19
CHAPITRE IV	
EXEMPLE À PARTIR DE DONNÉES SIMULÉES	29
4.1 Présentation des données	29
4.2 Cas non robuste - $\epsilon_i \sim \mathcal{N}(0, 1)$	32
4.3 Cas robuste - $\epsilon_i \sim \text{DL-Pareto}(\phi, \theta, \tau)$	34
CHAPITRE V	
ÉTUDE DE CAS	41
CONCLUSION	49
BIBLIOGRAPHIE	51

LISTE DES FIGURES

Figure	Page
3.1 Comportement d'une densité DL-Pareto(ϕ, θ, τ) lorsqu'un seul paramètre varie à la fois.	24
3.2 Comportement d'une densité DL-GenGamma($\gamma, \eta, \phi, \theta, \tau$) lorsqu'un seul paramètre varie à la fois (partie 1).	25
3.3 Comportement d'une densité DL-GenGamma($\gamma, \eta, \phi, \theta, \tau$) lorsqu'un seul paramètre varie à la fois (partie 2).	26
3.4 Comportement d'une densité DL-normale(ϕ, θ, τ) lorsqu'un seul paramètre varie à la fois.	27
3.5 Comportement d'une densité DL-GEP(0, 0, 1, 2, 2, 0.03).	28
4.1 Échantillon sans données aberrantes.	31
4.2 Densités <i>a posteriori</i> avec les données sans valeurs aberrantes et sous l'hypothèse de normalité.	32
4.3 Densités <i>a posteriori</i> avec les couples 1 à 9 d'origine et $(x_{10}, y_{10}) = (9.9, 990)$, sous l'hypothèse de normalité.	33
4.4 Densités <i>a posteriori</i> avec les données sans valeurs aberrantes.	35
4.5 Densités <i>a posteriori</i> avec les couples 1 à 9 d'origine et $(x_{10}, y_{10}) = (9.9, 990)$	36
4.6 Densités <i>a posteriori</i> avec les couples 1 à 8 d'origine, $(x_9, y_9) = (8.3, 830)$ et $(x_{10}, y_{10}) = (9.9, 990)$	37
4.7 Illustration de la convergence de mesures basées sur l'échantillon complet (avec une donnée aberrante) vers celles excluant la valeur aberrante.	38
4.8 Illustration de l'impact du choix de distribution sur l'inférence en présence d'une valeur aberrante.	39
5.1 Illustration du modèle de régression où les droites sont estimées par la médiane <i>a posteriori</i>	43
5.2 Illustration des densités <i>a posteriori</i> résultante.	44

5.3	Illustration du modèle de régression où les droites sont estimées par la médiane <i>a posteriori</i>	46
5.4	Illustration des densités <i>a posteriori</i> résultante.	47

LISTE DES TABLEAUX

Tableau	Page
4.1 Échantillon sans données aberrantes.	30
4.2 Mesures expliquant la variabilité des observations z dans l'échantillon sans données aberrantes.	30
5.1 Analyse descriptive des rendements quotidiens (%) du S&P/TSX et S&P 500 pour les dix dernières années.	41
5.2 Échelle de décision concernant la cote de Bayes.	45

RÉSUMÉ

La robustesse d'une inférence *a posteriori* face à des valeurs aberrantes est étudiée, dans un cadre de régression linéaire bayésienne. Il est montré que le modèle de régression linéaire simple avec une constante nulle peut être vu comme un problème d'inférence sur des paramètres de position-échelle. Un lien est alors effectué avec la théorie de robustesse de Desgagné (2011). En présence minoritaire de données aberrantes, une convergence en loi des densités *a posteriori* vers celles excluant les valeurs extrêmes, lorsque celles-ci tendent vers plus ou moins l'infini, est garantie sous une condition relative à des ailes suffisamment relevées de la densité des erreurs. Il est démontré que les estimations par maximum de vraisemblance sont eux aussi robustes. De plus, une nouvelle famille de densités, appelée DL-GEP, est proposée afin de guider l'utilisateur dans une recherche de distributions respectant le critère de robustesse. Les résultats théoriques sont illustrés d'abord à l'aide d'un exemple basé sur des données simulées, puis par une étude de cas s'appuyant sur des données financières, où les considérations pratiques sont abordées.

Mots-clés : Inférence bayésienne, robustesse, régression linéaire, valeurs aberrantes, paramètres de position-échelle, distributions à ailes relevées, fonctions à variation L-exponentielle, famille de densités DL-GEP.

INTRODUCTION

En statistique, et en particulier dans un cadre de régression linéaire, il arrive fréquemment qu'un échantillon de données contienne des valeurs aberrantes (des observations non représentatives ou extrêmes). Conséquemment, l'inférence statistique peut être contaminée par leur présence, menant ainsi à des résultats qui ne sont pas en accord avec la majorité des observations. La méthode de régression des moindres carrés est certes la plus populaire, mais elle peut mener à une inférence non robuste face à ces valeurs aberrantes. Il existe dans la littérature bayésienne plusieurs méthodes alternatives dites robustes. Cette robustesse peut être perçue sous trois angles différents.

Premièrement, le côté subjectif de l'information *a priori* incluse dans un modèle bayésien est souvent critiqué à l'effet qu'elle modifie la nature de l'analyse statistique. De plus, elle peut mener à une inférence non robuste lorsqu'il y a contradiction entre la distribution *a priori* et les données. Une solution possible est l'utilisation de lois *a priori* dites non informatives. Toutefois, dans certains contextes il peut être utile ou même nécessaire d'inclure de l'information *a priori*. C'est pourquoi plusieurs théories ont été développées afin de limiter son influence en cas de contradiction avec l'information véhiculée par les observations, voir Zhu, Ibrahim et Tang (2011) pour une revue de littérature à ce sujet. Cet aspect n'est pas étudié dans ce mémoire puisque des lois *a priori* non informatives sont utilisées.

Ensuite, il est fréquent que la fonction reliant les variables explicatives à la variable dépendante ne soit pas linéaire ou que les erreurs n'aient pas toutes la même variabilité. L'inférence découlant de la méthode de régression par moindres carrés peut alors être contaminée par des hypothèses du modèle non satisfaites. La régression non paramétrique est une approche permettant de remédier à ces problèmes de robustesse puisqu'elle ne se base habituellement pas sur les hypothèses classiques de régression

linéaire. Plusieurs auteurs ont proposé des solutions à cet égard en incluant aussi une certaine protection face aux valeurs aberrantes. Szpiro, Rice et Lumle (2010) suggèrent une méthode selon laquelle la variable dépendante est normale avec une espérance qui est fonction des variables explicatives et une variance évoluant avec la valeur des variables explicatives, ce qui assure une certaine robustesse face aux valeurs aberrantes. Dans une autre optique, Faul et Tipping (2001) modélisent la variable dépendante par une somme pondérée de fonctions des variables explicatives. De plus, les erreurs sont distribuées selon un mélange de deux densités, soit deux lois normales où la deuxième composante permet une plus grande variabilité, ou une loi normale et une loi uniforme. Barnes (2010) présente une approche semblable de mélange de lois en considérant plutôt une loi Student comme composante dédiée aux valeurs aberrantes. Mitra, Veeraraghavan et Chellappa (2010) utilisent aussi une méthode similaire où le bruit est la somme de deux composantes normales. L'un des deux termes permet une plus grande variabilité. Kuss (2006), quant à lui, compare quelques approches semblables aux précédentes. Il illustre, à l'aide d'exemples, la performance de trois alternatives où la différence réside dans la distribution du terme d'erreur. Plus précisément, il utilise une loi Student, une loi Laplace ou un mélange de deux lois normales. D'autre part, Yu et Moyeed (2001) modélisent un quantile de la variable dépendante plutôt que l'espérance, comme c'est habituellement le cas, et ce de manière non paramétrique. Une fonction de perte est choisie en conséquence. Elle est équivalente à supposer une distribution Laplace asymétrique sur le terme d'erreur. Lorsque la médiane est modélisée, la distribution devient une loi Laplace symétrique et celle-ci est reconnue pour avoir une plus grande résistance face aux valeurs aberrantes. Par ailleurs, Yang, Zhang et Sun (2007) proposent de restreindre le sous-espace des observations à celui contenant aucune valeur aberrante et la majorité des données. Les observations non contenues dans ce sous-espace sont exclues de la vraisemblance. Un algorithme permettant de trouver la forme de ce sous-espace est décrit.

Finalement, il y a l'angle traité dans ce mémoire, soit un cadre de régression linéaire paramétrique où l'objectif est l'obtention d'une inférence *a posteriori* robuste en présence

de valeurs aberrantes. Box et Tiao (1968) ont été les premiers auteurs à proposer une alternative robuste. Ils ont présenté l'idée de la modélisation des erreurs par un mélange de deux lois normales. La première est dédiée aux observations régulières et l'autre, permettant une plus grande variabilité, explique les valeurs extrêmes. Ces auteurs ont influencé la direction de plusieurs recherches subséquentes, que ce soit en régression robuste non paramétrique, tel que décrit plus haut, ou en régression paramétrique. En effet, West (1984) a généralisé ce principe en modélisant les erreurs par un mélange continu de lois normales afin d'obtenir une densité à ailes plus relevées. Liu (1996) a travaillé dans la même optique en utilisant lui aussi un mélange continu de lois normales. Dans cette situation, si le mélange a une distribution Gamma, alors la densité résultante est une loi Student. Fonseca, Ferreira et Migon (2008) modélisent directement les erreurs par une loi Student puisque celle-ci laisse place à des valeurs plus extrêmes qu'une loi normale. Aussi, ils développent une analyse bayésienne dite objective à l'aide de certaines lois *a priori* non informatives. Pour leur part, Peña, Zamar et Yan (2009) présentent une approche différente. Une région où se trouvent les valeurs non aberrantes est définie. Un poids de 1 est attribué à toutes les observations dans cette région. Pour celles qui sont à l'extérieur, un poids en fonction de la distance avec la majorité des données est imputé, celui-ci décroissant avec cette distance. Il est montré que la densité *a posteriori* résultante est près, au sens de la divergence de Kullback-Leibler, de celle excluant les valeurs aberrantes. De plus, il est démontré que la simple utilisation de densités à ailes relevées est insuffisante pour obtenir une inférence robuste, selon leur mesure stricte de robustesse.

Ce mémoire traite d'inférence *a posteriori* robuste en présence d'observations non représentatives de la variable dépendante. L'analyse se situe dans un cadre de régression linéaire simple où la constante est supposée nulle. Le modèle peut alors être vu comme un problème d'inférence sur des paramètres de position-échelle, ce qui permet, avec l'utilisation de lois *a priori* non informatives, d'appliquer les résultats de robustesse de Desgagné (2011). Le résultat principal de ce mémoire se résume à ceci : en présence minoritaire de valeurs aberrantes et sous une condition relative aux ailes de la distri-

bution des erreurs, une convergence en loi des densités *a posteriori* vers celles excluant les valeurs aberrantes, lorsque celles-ci tendent vers plus ou moins l'infini, est garantie. Cette condition requiert que la densité des erreurs ait des ailes suffisamment relevées. Ce résultat peut sembler en contradiction avec ce qu'affirment Peña, Zamar et Yan (2009), mais cela provient simplement d'une définition de robustesse différente, soit la convergence en loi versus la divergence de Kullback-Leibler. Cette dernière est basée sur un ratio de deux densités *a posteriori*, soit celle considérant l'échantillon complet et celle excluant les valeurs aberrantes. Cette mesure peut tendre vers l'infini lorsque ce rapport est évalué au niveau des ailes dû à une division de la forme 0/0. Cela ne pose pas de problème avec la convergence en loi, puisqu'elle est plutôt basée sur une différence de ces densités. Notons aussi que notre méthode se démarque par sa simplicité puisqu'il s'agit seulement de supposer que les erreurs ont une densité avec des ailes suffisamment relevées. Tout le reste de l'analyse se situe dans le cadre bayésien usuel.

Ce mémoire est organisé comme suit. Dans le chapitre 1, le contexte mathématique est décrit. Dans la section 1.1, le modèle de régression est exposé sous sa forme usuelle. Dans la section 1.2, il est présenté sous l'angle d'un problème d'inférence sur des paramètres de position-échelle. Ensuite, à la section 1.3, il est situé dans un contexte d'inférence robuste. Dans le chapitre 2, les résultats de robustesse de Desgagné (2011) sont adaptés au modèle de régression utilisé. Dans la section 2.1, des familles de fonctions caractérisant le comportement d'ailes sont définies. Elles permettront éventuellement d'établir une condition menant à la robustesse. Dans la section 2.2, l'impact de l'utilisation de densités à ailes suffisamment relevées est analysé. De plus, les résultats principaux sont présentés. Dans le chapitre 3, une nouvelle famille de densités, appelée DL-GEP, est proposée afin de guider l'utilisateur dans une recherche de distributions respectant le critère de robustesse. Dans le chapitre 4, les résultats sont illustrés à l'aide d'une analyse de données simulées. De plus, une étude de cas s'appuyant sur des données financières, où les considérations pratiques sont abordées, est développée dans le chapitre 5. Finalement, la conclusion est présentée à la suite du chapitre 5.

CHAPITRE I

CONTEXTE MATHÉMATIQUE

Dans ce chapitre, le modèle de régression utilisé tout au long du mémoire est présenté, tout d'abord sous sa forme usuelle puis comme un problème d'inférence sur des paramètres de position-échelle. Ensuite, il est situé dans un contexte de robustesse.

1.1 Présentation du modèle de régression sous sa forme usuelle

Le contexte de régression linéaire bayésienne, établi de manière à s'harmoniser avec les résultats de robustesse de Desgagné (2011), est le suivant.

- i) Soit n couples aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$. On suppose que le lien entre Y_i et X_i est le suivant :

$$Y_i = \beta X_i + \epsilon_i, \quad i = 1, \dots, n,$$

où $Y_i \in \mathbb{R}$ est la variable d'intérêt, $X_i \in \mathbb{R}$ est la variable explicative, $\beta \in \mathbb{R}$ est la pente de la droite de régression et $\epsilon_i \in \mathbb{R}$ est le terme d'erreur. Le modèle de régression linéaire simple avec une constante nulle est utilisé. Il peut alors être vu comme un problème d'inférence sur des paramètres de position-échelle (voir section 1.2). Ainsi, un lien peut être établi avec la théorie d'inférence robuste de Desgagné (2011). Généraliser au modèle de régression linéaire simple avec constante non-nulle n'est pas trivial.

- ii) Les variables aléatoires $\epsilon_1, \dots, \epsilon_n$ sont supposées conditionnellement indépendantes sachant σ , un paramètre d'échelle, avec une densité conditionnelle donnée par

$\epsilon_i \mid \sigma \sim (1/\sigma)f(\epsilon_i/\sigma)$, où f est continue, propre, strictement positive et symétrique ($f(a) = f(|a|)$, $\forall a \in \mathbb{R}$) avec des ailes monotones. La monotonie des ailes est définie, dans ce mémoire, de la manière suivante : il existe une constante $A > 0$ telle que $y > A$ implique que $f(y + \mu) \leq f(y)$, $\forall \mu > 0$. Ceci est l'équivalent, par symétrie de f , à ceci : $y < -A$ implique que $f(y - \mu) \leq f(y)$, $\forall \mu > 0$.

- iii) Les distributions *a priori* de β et σ sont respectivement proportionnelles à 1 et $1/\sigma$ avec $\beta \in \mathbb{R}$ et $\sigma > 0$. Il s'agit de densités impropres utilisées fréquemment comme lois non informatives. L'information repose alors presque entièrement sur les données, ce qui est en accord, sur ce point, avec l'approche fréquentiste. D'autre part, le paramètre β est supposé indépendant de $\epsilon_1, \dots, \epsilon_n$ et σ . La densité conjointe *a priori* de β et σ est alors donnée par $\pi(\beta, \sigma) \propto 1/\sigma$, $\beta \in \mathbb{R}$, $\sigma > 0$.
- iv) Les variables aléatoires X_1, \dots, X_n sont supposées indépendantes de β, σ et $\epsilon_1, \dots, \epsilon_n$ avec une densité quelconque.

1.2 Le modèle de régression présenté sous l'angle d'un problème de position-échelle

Dans cette section, le modèle de régression est présenté sous l'angle d'un problème d'inférence sur des paramètres de position-échelle. Ceci permettra d'appliquer les résultats de robustesse de Desgagné (2011).

Proposition 1. *Les variables aléatoires Y_1, \dots, Y_n sont conditionnellement indépendantes sachant $\mathbf{X} = (X_1, \dots, X_n)$, β et σ avec une densité conditionnelle donnée par $Y_i \mid X_i, \beta, \sigma \sim (1/\sigma)f((y_i - \beta x_i)/\sigma)$, $i = 1, \dots, n$.*

Ce résultat découle directement du changement de variables $Y_i = \beta X_i + \epsilon_i \Leftrightarrow \epsilon_i = Y_i - \beta X_i$ à partir de la densité des erreurs.

Proposition 2. *Les variables aléatoires $Z_1, \dots, Z_n = Y_1/X_1, \dots, Y_n/X_n$ sont conditionnellement indépendantes sachant \mathbf{X} , β et σ avec une densité conditionnelle donnée par*

$$Z_i \mid X_i, \beta, \sigma \sim \frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) = \frac{1}{\sigma} f_i\left(\frac{z_i - \beta}{\sigma}\right) \text{ où } f_i(a) = \frac{1}{1/|x_i|} f\left(\frac{a}{1/x_i}\right),$$

avec $a \in \mathbb{R}, i = 1, \dots, n$. En considérant les densités f_i , β joue le rôle d'un paramètre de position et σ d'un paramètre d'échelle. Ces densités ne diffèrent que par un paramètre d'échelle fixe $1/|x_i|$.

Ce résultat découle directement du changement de variables $Z_i = Y_i/X_i \Leftrightarrow Y_i = Z_i X_i$ à partir de la densité de la variable aléatoire Y_i .

Proposition 3. *L'inférence a posteriori basée sur les données \mathbf{X} et $\mathbf{Y} = (Y_1, \dots, Y_n)$ est la même que celle basée sur \mathbf{X} et $\mathbf{Z} = (Z_1, \dots, Z_n)$, c'est-à-dire que $\pi(\beta, \sigma | \mathbf{y}, \mathbf{x}) = \pi(\beta, \sigma | \mathbf{z}, \mathbf{x})$, où $\pi(\beta, \sigma | \cdot, \mathbf{x})$ représente la densité conjointe a posteriori de β et σ .*

Démonstration. D'abord, on note que la densité conjointe de \mathbf{Y} et \mathbf{X} sachant β et σ peut s'écrire comme

$$\begin{aligned} f_{\mathbf{Y}, \mathbf{X}}(\mathbf{y}, \mathbf{x} | \beta, \sigma) &= f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}, \beta, \sigma) f_{\mathbf{X}}(\mathbf{x} | \beta, \sigma) \\ &= \prod_{i=1}^n f_{Y_i|X_i}(y_i | x_i, \beta, \sigma) f_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

De la même manière,

$$\begin{aligned} f_{\mathbf{Z}, \mathbf{X}}(\mathbf{z}, \mathbf{x} | \beta, \sigma) &= f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} | \mathbf{x}, \beta, \sigma) f_{\mathbf{X}}(\mathbf{x} | \beta, \sigma) \\ &= \prod_{i=1}^n f_{Z_i|X_i}(z_i | x_i, \beta, \sigma) f_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

Alors, en posant $z_i = y_i/x_i$,

$$\begin{aligned} \pi(\beta, \sigma | \mathbf{y}, \mathbf{x}) / \pi(\beta, \sigma) &\propto f_{\mathbf{Y}, \mathbf{X}}(\mathbf{y}, \mathbf{x} | \beta, \sigma) \propto \prod_{i=1}^n f_{Y_i|X_i}(y_i | x_i, \beta, \sigma) \\ &= \prod_{i=1}^n (1/\sigma) f((y_i - \beta x_i)/\sigma) \\ &= \prod_{i=1}^n (1/\sigma) f((z_i - \beta)/(\sigma/x_i)) \\ &\propto \prod_{i=1}^n f_{Z_i|X_i}(z_i | x_i, \beta, \sigma) \propto f_{\mathbf{Z}, \mathbf{X}}(\mathbf{z}, \mathbf{x} | \beta, \sigma) \\ &\propto \pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) / \pi(\beta, \sigma), \end{aligned}$$

$$\Rightarrow \pi(\beta, \sigma | \mathbf{y}, \mathbf{x}) = \pi(\beta, \sigma | \mathbf{z}, \mathbf{x}).$$

■

Ce sont alors les variables aléatoires Z_i qui seront considérées afin d'obtenir les résultats d'inférence robuste. En effet, vu de cet angle, le contexte devient compatible avec celui présenté dans Desgagné (2011).

1.3 Le modèle de régression dans un contexte d'inférence robuste

L'inférence sur β et σ en présence d'observations non représentatives de la variable d'intérêt est étudiée. Étant donné que les densités *a priori* de β et σ sont impropres et dites non-informatives, il ne peut y avoir de conflit entre l'information *a priori* et les observations. Par ailleurs, on suppose que l'utilisateur a un contrôle sur la variable explicative et par conséquent qu'il n'y a aucune valeur aberrante parmi les observations x . Par exemple, dans le domaine de l'assurance automobile, on étudie l'effet de l'âge d'un assuré sur son montant annuel de réclamations. Il est alors raisonnable de supposer que les âges observés sont dépourvus de valeurs aberrantes.

Les résultats de ce mémoire sont de nature asymptotique. Cependant, ils ne le sont pas au sens habituel, lorsqu'on considère que le nombre d'observations tend vers l'infini, mais plutôt au sens où on considère que les valeurs extrêmes des z_i tendent vers plus ou moins l'infini. Ceci est noté, d'une part, par $z_i \rightarrow \infty$ si z_i est une valeur aberrante positive, ce qui est équivalent à soit $y_i \rightarrow \infty, x_i > 0$ ou $y_i \rightarrow -\infty, x_i < 0$. D'autre part, c'est noté par $z_i \rightarrow -\infty$ si z_i est une valeur aberrante négative, ce qui est équivalent à soit $y_i \rightarrow \infty, x_i < 0$ ou $y_i \rightarrow -\infty, x_i > 0$. On exclut la possibilité que $z_i = y_i/x_i \rightarrow \pm\infty$ dans le cas où $x_i = 0$, puisque X_i est une variable aléatoire continue et que $X_i = 0$ est de probabilité nulle. En pratique, si x_i est très près de 0, on aura toutefois l'impression que $z_i = y_i/x_i$ est une valeur aberrante. Il faut toutefois relativiser puisque le paramètre d'échelle $\sigma/|x_i|$ de Z_i augmentera aussi avec une petite valeur de $|x_i|$. De plus, $z_i/(\sigma/|x_i|) = y_i/\sigma$, la nature aberrante de z_i ne dépend donc pas de x_i , mais plutôt de y_i .

Parmi les n observations, notées par $\mathbf{z} = (z_1, \dots, z_n)$, on suppose que k d'entre elles, notées par \mathbf{z}_k , forment un groupe de données non aberrantes (ou valeurs fixes). On suppose que l d'entre elles, notées par \mathbf{z}_l , sont considérées aberrantes à gauche par rapport

aux données fixes et que r d'entre elles, notées par \mathbf{z}_r , sont considérées aberrantes à droite par rapport aux données fixes avec $0 \leq k, l, r \leq n$ et $k + l + r = n$.

Afin d'identifier la nature de z_i , trois fonctions binaires sont définies, soit k_i, l_i et r_i , qui sont égales à 1 si z_i est respectivement membre de $\mathbf{z}_k, \mathbf{z}_l$ ou \mathbf{z}_r , et 0 sinon, avec $k_i + l_i + r_i = 1$ pour chaque z_i , et $\sum_{i=1}^n k_i = k$, $\sum_{i=1}^n l_i = l$ et $\sum_{i=1}^n r_i = r$. On définit $\mathbf{x}_k, \mathbf{x}_l$ et \mathbf{x}_r comme les vecteurs contenant respectivement tous les x_i tels que $z_i \in \mathbf{z}_k$, tous les x_i tels que $z_i \in \mathbf{z}_l$ et tous les x_i tels que $z_i \in \mathbf{z}_r$.

La densité *a posteriori* étudiée est donnée par

$$\pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) = (m(\mathbf{z} | \mathbf{x}))^{-1} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right),$$

avec

$$m(\mathbf{z} | \mathbf{x}) = \int_{\beta} \int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) d\sigma d\beta.$$

Aussi, on définit

$$\pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k) = (m(\mathbf{z}_k | \mathbf{x}_k))^{-1} \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) \right]^{k_i},$$

avec

$$m(\mathbf{z}_k | \mathbf{x}_k) = \int_{\beta} \int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) \right]^{k_i} d\sigma d\beta.$$

La densité $\pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$ correspond à la densité *a posteriori* tenant compte que des observations $(\mathbf{z}_k, \mathbf{x}_k)$ avec une densité *a priori* conjointe de β et σ proportionnelle à $1/\sigma$. De manière analogue, les quantités suivantes sont définies :

$$\begin{aligned} \pi(\beta, \sigma | \mathbf{z}_l, \mathbf{x}_l) &\propto \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) \right]^{l_i}, \\ \pi(\beta, \sigma | \mathbf{z}_r, \mathbf{x}_r) &\propto \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma/|x_i|} f\left(\frac{z_i - \beta}{\sigma/x_i}\right) \right]^{r_i}. \end{aligned}$$

Les densités $\pi(\beta, \sigma | \mathbf{z}_l, \mathbf{x}_l)$ et $\pi(\beta, \sigma | \mathbf{z}_r, \mathbf{x}_r)$ correspondent respectivement à la densité *a posteriori* tenant compte que des observations $(\mathbf{z}_l, \mathbf{x}_l)$ et à celle tenant compte que des observations $(\mathbf{z}_r, \mathbf{x}_r)$, avec une densité *a priori* proportionnelle à $1/\sigma$. On note que

$$\sigma\pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) \propto \sigma\pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k) \times \sigma\pi(\beta, \sigma | \mathbf{z}_l, \mathbf{x}_l) \times \sigma\pi(\beta, \sigma | \mathbf{z}_r, \mathbf{x}_r).$$

En effet, la vraisemblance de (\mathbf{z}, \mathbf{x}) , proportionnelle à $\sigma\pi(\beta, \sigma | \mathbf{z}, \mathbf{x})$, est le produit des vraisemblances des trois groupes d'observations $(\mathbf{z}_k, \mathbf{x}_k)$, $(\mathbf{z}_l, \mathbf{x}_l)$ et $(\mathbf{z}_r, \mathbf{x}_r)$.

La densité *a posteriori* considérant les observations (\mathbf{y}, \mathbf{x}) est donnée par

$$\pi(\beta, \sigma | \mathbf{y}, \mathbf{x}) = (m(\mathbf{y} | \mathbf{x}))^{-1} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma} f\left(\frac{y_i - \beta x_i}{\sigma}\right),$$

avec

$$m(\mathbf{y} | \mathbf{x}) = \int_{\beta} \int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma} f\left(\frac{y_i - \beta x_i}{\sigma}\right) d\sigma d\beta.$$

Aussi, on définit

$$\pi(\beta, \sigma | \mathbf{y}_k, \mathbf{x}_k) = (m(\mathbf{y}_k | \mathbf{x}_k))^{-1} \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma} f\left(\frac{y_i - \beta x_i}{\sigma}\right) \right]^{k_i},$$

avec

$$m(\mathbf{y}_k | \mathbf{x}_k) = \int_{\beta} \int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \left[\frac{1}{\sigma} f\left(\frac{y_i - \beta x_i}{\sigma}\right) \right]^{k_i} d\sigma d\beta.$$

Dans ce mémoire, l'objectif est d'assurer une inférence robuste en présence de données aberrantes, ce qui se traduit par une certaine proximité des densités $\pi(\beta, \sigma | \mathbf{z}, \mathbf{x})$ et $\pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$ lorsque les valeurs extrêmes tendent vers plus ou moins l'infini. Cet aspect est approfondi dans le prochain chapitre.

CHAPITRE II

INFÉRENCE ROBUSTE

Le théorème de robustesse est présenté à la section 2.2, mais d'abord on présente à la section 2.1 une nouvelle famille de fonctions, dites à variation L-exponentielle, nécessaire à la vérification d'une condition du théorème.

2.1 Famille de fonctions à variation L-exponentielle

D'abord, quelques définitions présentées par Desgagné (2011), portant sur le comportement de l'aile d'une fonction, sont décrites. Elles sont nécessaires afin de quantifier l'épaisseur des ailes d'une densité. Ainsi, il pourra être vérifié que la distribution du terme d'erreur du modèle de régression engendre une inférence robuste.

Desgagné (2011) a d'abord généralisé les fonctions à variation lente et régulière à une classe de fonctions à variation exponentielle. Rappelons qu'une fonction g à variation lente est telle que $g(\nu y)/g(y) \rightarrow 1$ lorsque $y \rightarrow \infty, \forall \nu > 0$, et une fonction g à variation régulière indexée par $\rho \in \mathbb{R}$ est telle que $g(\nu y)/g(y) \rightarrow \nu^\rho$ lorsque $y \rightarrow \infty, \forall \nu > 0$. La classe de fonctions à variation exponentielle est définie comme suit.

Définition 1. Une fonction g est dite à variation exponentielle à ∞ indexée par (γ, δ, α) , notée $g \in E_{\gamma, \delta, \alpha}(\infty)$, si pour $y > 0$, g peut être écrite comme

$$g(y) = e^{-\delta y^\gamma} y^{-\alpha} S(y),$$

où $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}$, et $S(y)$ est une fonction à variation lente.

Ces fonctions sont généralisées à la classe de fonctions à variation L-exponentielle par un changement de variable $y = \log z$ et $h(z) = g(\log z)$. Le terme L (pour log) représente la transformation logarithmique/exponentielle.

Définition 2. Une fonction h est dite à variation L-lente à ∞ , notée $h \in L_{0,0,0}(\infty)$, si pour $\nu > 0$,

$$h(z^\nu)/h(z) \rightarrow 1 \text{ lorsque } z \rightarrow \infty.$$

Par exemple, $h(z) = \log(\log z) \in L_{0,0,0}(\infty)$ ou $h(z) = 1 \in L_{0,0,0}(\infty)$.

Définition 3. Une fonction h est dite à variation L-régulière à ∞ indexée par $\rho \in \mathbb{R}$, notée $h \in L_{0,0,-\rho}(\infty)$, si pour $\nu > 0$,

$$h(z^\nu)/h(z) \rightarrow \nu^\rho \text{ lorsque } z \rightarrow \infty,$$

ou de manière équivalente, si pour $z > 1$, h peut être écrite comme

$$h(z) = (\log z)^\rho T(z),$$

avec $T \in L_{0,0,0}(\infty)$.

Définition 4. Une fonction h est dite à variation L-exponentielle à ∞ indexée par (γ, δ, α) , notée $h \in L_{\gamma,\delta,\alpha}(\infty)$, si pour $z > 1$, h peut être écrite comme

$$h(z) = e^{-\delta(\log z)^\gamma} (\log z)^{-\alpha} T(z),$$

avec $T \in L_{0,0,0}(\infty)$, $\gamma \geq 0$, $\delta \geq 0$, $\alpha \in \mathbb{R}$.

Par convention, $\gamma = 0$ si et seulement si $\delta = 0$. Une fonction à variation L-exponentielle est à variation L-lente si $\gamma = 0$, $\delta = 0$, $\alpha = 0$ et elle est à variation L-régulière indexée par ρ si $\gamma = 0$, $\delta = 0$, $\alpha = -\rho$. D'autre part, on peut montrer facilement que si $h_1(z)/h_2(z) \rightarrow 1$ lorsque $z \rightarrow \infty$ et $h_1 \in L_{\gamma,\delta,\alpha}(\infty)$, alors $h_2 \in L_{\gamma,\delta,\alpha}(\infty)$.

Finalement, voici une proposition qui servira à simplifier une condition du théorème présenté dans la prochaine section.

Proposition 4. *Si $h \in L_{\gamma, \delta, \alpha}(\infty)$ avec $\gamma < 1$, alors $\forall \tau > 0$,*

$$h(\tau z)/h(z) \rightarrow 1 \text{ lorsque } z \rightarrow \infty.$$

La preuve de ce résultat se trouve dans Desgagné (2011).

2.2 Théorème de robustesse

Le théorème principal de ce mémoire est maintenant présenté selon le contexte bayésien décrit au chapitre 1. Il s'agit d'une adaptation des résultats de Desgagné (2011) au contexte de régression. L'écriture $\omega \rightarrow \infty$, où $\omega = \min(-z_l, z_r)$, signifie que chaque terme du vecteur $(-z_l, z_r)$ tend vers l'infini à n'importe quelle vitesse donnée.

Théorème 1. *Si les conditions suivantes sont satisfaites :*

i) $zf(z) \in L_{\gamma, \delta, \alpha}(\infty)$ avec $\gamma < 1$,

ii) $k - (l + r) \geq 2$,

alors nous obtenons les résultats suivants :

a) $m(\mathbf{z} | \mathbf{x}) / \left(m(\mathbf{z}_k | \mathbf{x}_k) \prod_{i=1}^n [|x_i| f(x_i z_i)]^{l_i + r_i} \right) \rightarrow 1$ lorsque $\omega \rightarrow \infty$,

b) $\pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) \rightarrow \pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$, $\beta \in \mathbb{R}, \sigma > 0$ lorsque $\omega \rightarrow \infty$,

c) $\sigma \pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) \rightarrow 0$ lorsque $\beta \rightarrow \pm\infty$ et/ou $\sigma \rightarrow \infty$ ou $\sigma \rightarrow 0$, $\omega \rightarrow \infty$,

d) $\beta, \sigma | \mathbf{z}, \mathbf{x} \xrightarrow{\mathcal{L}} \beta, \sigma | \mathbf{z}_k, \mathbf{x}_k$ lorsque $\omega \rightarrow \infty$,

e) $\beta | \mathbf{z}, \mathbf{x} \xrightarrow{\mathcal{L}} \beta | \mathbf{z}_k, \mathbf{x}_k$ lorsque $\omega \rightarrow \infty$,

f) $\sigma | \mathbf{z}, \mathbf{x} \xrightarrow{\mathcal{L}} \sigma | \mathbf{z}_k, \mathbf{x}_k$ lorsque $\omega \rightarrow \infty$,

g) $\arg \max_{\beta, \sigma} \mathcal{L}(\beta, \sigma | \mathbf{z}, \mathbf{x}) \rightarrow \arg \max_{\beta, \sigma} \mathcal{L}(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$ lorsque $\omega \rightarrow \infty$, où $\mathcal{L}(\cdot)$ est la fonction de vraisemblance usuelle.

Notons toutefois que dans Desgagné (2011), la condition i) est plutôt donnée par $zf_i(z) \in L_{\gamma, \delta, \alpha}(\infty)$ avec $\gamma < 1$. Mais, puisque $zf_i(z) = |x_i| zf(x_i z)$, alors $zf_i(z) \in L_{\gamma, \delta, \alpha}(\infty)$ avec $\gamma < 1$ est équivalent à $zf(z) \in L_{\gamma, \delta, \alpha}(\infty)$ avec $\gamma < 1$ selon la proposition 4 en posant $\tau = |x_i|$ et $h(z) = zf(z)$. La condition ii) requiert qu'il y ait au moins

deux valeurs fixes (k) de plus que de données aberrantes ($l+r$). Il s'agit d'une condition raisonnable puisqu'on s'attend généralement à ce que les valeurs aberrantes ne soient pas la norme et on a besoin d'au moins deux observations « nettes » pour estimer à la fois la position et l'échelle. Le résultat a) décrit le comportement asymptotique de la densité marginale de \mathbf{Z} considérant \mathbf{X} fixe, lorsque $\omega \rightarrow \infty$. On peut montrer que ce résultat est aussi équivalent à

$$m(\mathbf{y} | \mathbf{x}) / \left(m(\mathbf{y}_k | \mathbf{x}_k) \prod_{i=1}^n [f(y_i)]^{l_i+r_i} \right) \rightarrow 1 \text{ lorsque } \omega \rightarrow \infty.$$

Le résultat b) indique le comportement asymptotique de la densité *a posteriori* conjointe de β et σ , pour n'importe quelles valeurs fixes de β et σ . La densité *a posteriori* de β et σ tend vers celle qui exclut les valeurs aberrantes. Notons que les résultats b) à g) sont aussi valables lorsque \mathbf{z} et \mathbf{z}_k sont remplacés par \mathbf{y} et \mathbf{y}_k puisque selon la proposition 3, les densités *a posteriori* et la vraisemblance sont les mêmes que l'on considère \mathbf{z} ou \mathbf{y} . Le résultat c) indique que la densité conjointe *a posteriori* de β et σ tend vers 0 lorsque $\beta \rightarrow \pm\infty$ et/ou $\sigma \rightarrow \infty$ ou $\sigma \rightarrow 0$, $\omega \rightarrow \infty$. Ceci implique qu'elle tend vers 0 dans la région des valeurs aberrantes. Il pourrait y avoir un mode dans ces régions, mais sa hauteur tendrait alors vers 0. Le ratio $\pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) / \pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$ tend alors vers une forme 0/0 et sa limite peut prendre des valeurs entre 0 et ∞ , dépendant du lien entre ω , β et σ . Pour les résultats d), e) et f), la convergence en loi signifie respectivement que $\Pr(\beta \leq c, \sigma \leq d | \mathbf{z}, \mathbf{x}) \rightarrow \Pr(\beta \leq c, \sigma \leq d | \mathbf{z}_k, \mathbf{x}_k)$, $\Pr(\beta \leq c | \mathbf{z}, \mathbf{x}) \rightarrow \Pr(\beta \leq c | \mathbf{z}_k, \mathbf{x}_k)$ et $\Pr(\sigma \leq d | \mathbf{z}, \mathbf{x}) \rightarrow \Pr(\sigma \leq d | \mathbf{z}_k, \mathbf{x}_k)$ lorsque $\omega \rightarrow \infty$, $\forall c \in \mathbb{R}$ et $\forall d > 0$. Le résultat g) indique que sous l'approche fréquentiste, les estimations de β et σ par maximum de vraisemblance sont robustes en présence de valeurs aberrantes.

La section est complétée avec les démonstrations des résultats e) à g). Les résultats a) à d) du théorème 1 ne sont pas démontrés puisqu'ils proviennent directement de Desgagné (2011). En effet, comme il a été montré dans la proposition 3,

$$\pi(\beta, \sigma | \mathbf{y}, \mathbf{x}) = \pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) \propto (1/\sigma) \prod_{i=1}^n f_{Z_i|X_i}(z_i | x_i, \beta, \sigma).$$

Les variables aléatoires Z_i conditionnelles à X_i , β et σ peuvent alors être utilisées dans le calcul des densités *a posteriori*. Puis, comme la proposition 2 l'indique, β et σ jouent

respectivement le rôle d'un paramètre de position et d'échelle pour les variables aléatoires conditionnelles $Z_i \mid X_i, \beta, \sigma$, lorsque les densités f_i sont considérées. Alors, la théorie de Desgagné (2011) s'applique et les résultats a) à d) sont directement obtenus.

Démonstration du résultat e). Nous démontrons que le résultat e) découle du résultat d). Ce dernier peut s'écrire plus explicitement comme suit : $\forall \epsilon > 0$ il existe une constante $M(\epsilon) > 0$ telle que

$$\omega > M(\epsilon) \Rightarrow |\Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}, \mathbf{x}) - \Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}_k, \mathbf{x}_k)| < \epsilon, \forall c \in \mathbb{R}, \forall d > 0,$$

ou sous une forme alternative,

$$\omega > M(\epsilon) \Rightarrow |\Pr(\beta \leq c, \sigma > d \mid \mathbf{z}, \mathbf{x}) - \Pr(\beta \leq c, \sigma > d \mid \mathbf{z}_k, \mathbf{x}_k)| < \epsilon, \forall c \in \mathbb{R}, \forall d > 0.$$

Alors, il existe une constante $A(\epsilon) = M(\epsilon/2)$ telle que $\omega > A(\epsilon)$ implique que

$$\begin{aligned} & |\Pr(\beta \leq c \mid \mathbf{z}, \mathbf{x}) - \Pr(\beta \leq c \mid \mathbf{z}_k, \mathbf{x}_k)| = \\ & \quad |\Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}, \mathbf{x}) + \Pr(\beta \leq c, \sigma > d \mid \mathbf{z}, \mathbf{x}) \\ & \quad - \Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}_k, \mathbf{x}_k) - \Pr(\beta \leq c, \sigma > d \mid \mathbf{z}_k, \mathbf{x}_k)| \\ & \leq |\Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}, \mathbf{x}) - \Pr(\beta \leq c, \sigma \leq d \mid \mathbf{z}_k, \mathbf{x}_k)| \\ & \quad + |\Pr(\beta \leq c, \sigma > d \mid \mathbf{z}, \mathbf{x}) - \Pr(\beta \leq c, \sigma > d \mid \mathbf{z}_k, \mathbf{x}_k)| < \epsilon/2 + \epsilon/2 = \epsilon, \end{aligned}$$

$$\forall c \in \mathbb{R}, \forall d > 0. \quad \blacksquare$$

La démonstration du résultat f) est similaire.

Le résultat d) indique que tous les quantiles de la densité conjointe *a posteriori* convergent vers les quantiles de la densité excluant les valeurs aberrantes. Il s'agit d'un simple corollaire de la convergence en loi. Les résultats e) et f) indiquent alors que tous les quantiles des densités $\pi(\beta \mid \mathbf{z}, \mathbf{x})$ et $\pi(\sigma \mid \mathbf{z}, \mathbf{x})$ convergent respectivement vers ceux des densités $\pi(\beta \mid \mathbf{z}_k, \mathbf{x}_k)$ et $\pi(\sigma \mid \mathbf{z}_k, \mathbf{x}_k)$. Les estimations des paramètres β et σ par la médiane et l'intervalle de crédibilité à plus haute densité *a posteriori*, basés sur $\pi(\beta \mid \mathbf{z}, \mathbf{x})$ et $\pi(\sigma \mid \mathbf{z}, \mathbf{x})$, sont donc robustes face à des valeurs aberrantes.

Démonstration du résultat g). Le résultat b) indique que

$$\pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) \rightarrow \pi(\beta, \sigma \mid \mathbf{z}_k, \mathbf{x}_k), \beta \in \mathbb{R}, \sigma > 0 \text{ lorsque } \omega \rightarrow \infty.$$

Ce résultat peut s'écrire plus explicitement comme suit : $\forall \epsilon > 0, \forall \lambda > 0, \forall \tau > 1$, il existe une constante $M_1(\epsilon, \lambda, \tau) > 0$ telle que

$$\omega > M_1(\epsilon, \lambda, \tau), -\lambda < \beta < \lambda, 1/\tau < \sigma < \tau \Rightarrow |\pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) - \pi(\beta, \sigma \mid \mathbf{z}_k, \mathbf{x}_k)| < \epsilon.$$

Il en découle un corollaire, qu'on nomme le résultat b'), donné par : $\forall \epsilon > 0, \forall \lambda > 0, \forall \tau > 1$, il existe une constante $M_2(\epsilon, \lambda, \tau) > 0$ telle que

$$\omega > M_2(\epsilon, \lambda, \tau), -\lambda < \beta < \lambda, 1/\tau < \sigma < \tau \Rightarrow |\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) - \sigma \pi(\beta, \sigma \mid \mathbf{z}_k, \mathbf{x}_k)| < \epsilon.$$

Puisque $1/\sigma < \tau$, on n'a qu'à poser $M_2(\epsilon, \lambda, \tau) = M_1(\epsilon/\tau, \lambda, \tau)$ pour vérifier la validité du corollaire. On note que $\mathcal{L}(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) \propto \sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x})$.

Le résultat c) indique que $\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) \rightarrow 0$ lorsque $\beta \rightarrow \pm\infty$ et/ou $\sigma \rightarrow 0$ ou $\sigma \rightarrow \infty$, $\omega \rightarrow \infty$. Ce résultat peut s'écrire plus explicitement comme suit : $\forall \epsilon > 0, \exists M_3(\epsilon) > 1$ telle que $\omega > M_3(\epsilon)$ et $|\beta| > M_3(\epsilon)$ et/ou $|\log \sigma| > \log M_3(\epsilon)$ implique que

$$\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) < \epsilon.$$

On note $(\hat{\beta}_k, \hat{\sigma}_k)$ le mode de $\sigma \pi(\beta, \sigma \mid \mathbf{z}_k, \mathbf{x}_k)$, et $(\hat{\beta}_n, \hat{\sigma}_n)$ le mode de $\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x})$. On suppose que $\sigma \pi(\beta, \sigma \mid \mathbf{z}_k, \mathbf{x}_k)$ est unimodale. L'objectif est donc de montrer que $(\hat{\beta}_n, \hat{\sigma}_n)$ est unique et que $(\hat{\beta}_n, \hat{\sigma}_n) \rightarrow (\hat{\beta}_k, \hat{\sigma}_k)$ lorsque $\omega \rightarrow \infty$.

D'abord, on montre qu'il n'y a pas de mode dans les ailes de $\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x})$. Si on définit

$$\epsilon_0 = \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k \mid \mathbf{z}_k, \mathbf{x}_k)/10,$$

alors selon le résultat c), $\omega > M_3(\epsilon_0)$ et $|\beta| > M_3(\epsilon_0)$ et/ou $|\log \sigma| > \log M_3(\epsilon_0)$ implique que

$$\sigma \pi(\beta, \sigma \mid \mathbf{z}, \mathbf{x}) \leq \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k \mid \mathbf{z}_k, \mathbf{x}_k)/10.$$

On pose $\lambda = \max(M_3(\epsilon_0), |\hat{\beta}_k|)$ et $\tau = \max(M_3(\epsilon_0), \exp |\log \hat{\sigma}_k|)$ et

$$\mathcal{D} = [-\lambda, \lambda] \times [1/\tau, \tau].$$

On a alors que $(\hat{\beta}_k, \hat{\sigma}_k) \in \mathcal{D}$ et on peut restreindre la recherche de $(\hat{\beta}_n, \hat{\sigma}_n)$ à la région finie \mathcal{D} , puisque

$$\hat{\sigma}_n \pi(\hat{\beta}_n, \hat{\sigma}_n | \mathbf{z}, \mathbf{x}) > \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) / 10.$$

En effet, selon le résultat b'), on a $\forall \epsilon > 0$, $(\beta, \sigma) \in \mathcal{D}$ et $\omega > M_2(\epsilon, \lambda, \tau)$ implique que

$$|\sigma \pi(\beta, \sigma | \mathbf{z}, \mathbf{x}) - \sigma \pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)| < \epsilon,$$

et donc en particulier

$$-\epsilon < \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) - \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}, \mathbf{x}) < \epsilon. \quad (2.1)$$

On a donc que

$$\begin{aligned} \hat{\sigma}_n \pi(\hat{\beta}_n, \hat{\sigma}_n | \mathbf{z}, \mathbf{x}) &\geq \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}, \mathbf{x}) \\ &> \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) - \epsilon \\ &> \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) / 10, \end{aligned}$$

en choisissant $\epsilon < (9/10) \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k)$.

Prenons un couple $(\beta_0, \sigma_0) \in \mathcal{D}$ tel que $(\beta_0, \sigma_0) \neq (\hat{\beta}_k, \hat{\sigma}_k)$. Soit

$$d(\beta_0, \sigma_0) = \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) - \sigma_0 \pi(\beta_0, \sigma_0 | \mathbf{z}_k, \mathbf{x}_k),$$

où $d(\beta_0, \sigma_0) > 0$ puisqu'on suppose que $\sigma \pi(\beta, \sigma | \mathbf{z}_k, \mathbf{x}_k)$ est unimodale. Alors, si on choisit $\epsilon < d(\beta_0, \sigma_0)/4$ et on considère $\omega > M_2(\epsilon, \lambda, \tau)$, on en déduit par le résultat b') que

$$\begin{aligned} \sigma_0 \pi(\beta_0, \sigma_0 | \mathbf{z}, \mathbf{x}) &< \sigma_0 \pi(\beta_0, \sigma_0 | \mathbf{z}_k, \mathbf{x}_k) + \epsilon \\ &= \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}_k, \mathbf{x}_k) - d(\beta_0, \sigma_0) + \epsilon \\ &< \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}, \mathbf{x}) - d(\beta_0, \sigma_0) + 2\epsilon \\ &< \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}, \mathbf{x}) - d(\beta_0, \sigma_0) / 2. \end{aligned}$$

On a bien que $\sigma_0 \pi(\beta_0, \sigma_0 | \mathbf{z}, \mathbf{x}) < \hat{\sigma}_k \pi(\hat{\beta}_k, \hat{\sigma}_k | \mathbf{z}, \mathbf{x})$ et on a choisi (β_0, σ_0) arbitrairement. Alors, cela démontre que $(\hat{\beta}_n, \hat{\sigma}_n)$ est unique et qu'il est en $(\hat{\beta}_k, \hat{\sigma}_k)$ lorsque $\omega \rightarrow \infty$. Aussi, par l'équation (2.1), on voit que la valeur du mode convergera. ■

CHAPITRE III

FAMILLE DE DENSITÉS À VARIATION L-EXPONENTIELLE

Dans ce chapitre, une nouvelle famille de densités symétriques définie sur les réels, appelée DL-GEP, est présentée. Un choix approprié de ses paramètres permet l'obtention de distributions respectant le critère de robustesse. Elle est construite à partir de l'aile d'une distribution de puissance d'exponentielles généralisée (GEP), voir Desgagné et Angers (2007). Afin d'obtenir la DL-GEP, une transformation logarithmique/exponentielle est d'abord effectuée. Elle est suivie d'un changement de position vers l'origine et d'un passage des positifs aux réels. Ces transformations expliquent le terme *DL* placé devant le nom de la densité d'origine, où *D* (pour double) représente le passage des positifs aux réels et *L* (pour log) représente la transformation logarithmique/exponentielle.

Nous définissons l'aile normalisée de la densité GEP, notée $GEP^+(\gamma, \delta, \alpha, \beta, \theta)$, comme suit.

Définition 5. Si $V \sim GEP^+(\gamma, \delta, \alpha, \beta, \theta)$, alors sa densité est donnée par

$$g(v) = K(\gamma, \delta, \alpha, \beta, \theta) \cdot e^{-\delta v^\gamma} v^{-\alpha} (\log v)^{-\beta}, v \geq \theta,$$

où $K(\gamma, \delta, \alpha, \beta, \theta) = \left(\int_{\theta}^{\infty} e^{-\delta v^\gamma} v^{-\alpha} (\log v)^{-\beta} dv \right)^{-1}$ est la constante de normalisation, $\gamma \geq 0, \delta \geq 0, \alpha \in \mathbb{R}, \beta \in \mathbb{R}$. Par convention, $\gamma = 0$ si et seulement si $\delta = 0$. De plus,

- i) si $\beta \neq 0$, alors $\theta > 1$,
- ii) si $\beta = 0, \alpha \neq 0$, alors $\theta > 0$,
- iii) si $\beta = 0, \alpha = 0$, alors $\theta \geq 0$,

iv) si $\gamma = \delta = 0$ alors $\alpha > 1$ ou $\alpha = 1, \beta > 1$.

Les contraintes i) à iii) assurent que la densité est positive et bornée, alors que la contrainte iv) fait en sorte que la densité soit propre.

À partir de la variable aléatoire $V \sim \text{GEP}^+(\gamma, \delta, \alpha, \beta, \theta)$, les changements suivants sont apportés :

- transformation logarithmique/exponentielle ($W = e^V, W > e^\theta$),
- changement de position vers l'origine ($T = W - e^\theta, T > 0$),
- passage des positifs aux réels et ajout d'un paramètre d'échelle τ ($|Y| = \tau \times T, Y \in \mathbb{R}$).

Globalement, le changement de variable est $|Y| = \tau (e^V - e^\theta)$, ce qui est équivalent à $V = \log(|Y/\tau| + e^\theta)$. Le jacobien résultant est $(2\tau(|Y/\tau| + e^\theta))^{-1}$.

Définition 6. Une variable aléatoire Y a une distribution DL-GEP, notée $Y \sim \text{DL-GEP}(\gamma, \delta, \alpha, \beta, \theta, \tau)$, si sa densité est donnée par

$$f(y) = K(\gamma, \delta, \alpha, \beta, \theta) \cdot (2\tau(|y/\tau| + e^\theta))^{-1} \exp[-\delta(\log(|y/\tau| + e^\theta))^\gamma] \\ \times (\log(|y/\tau| + e^\theta))^{-\alpha} (\log(\log(|y/\tau| + e^\theta)))^{-\beta},$$

avec $y \in \mathbb{R}, \tau > 0$ et les mêmes contraintes que celles données à la définition 5 pour les paramètres de la densité.

Il peut être facilement vérifié que si $Y \sim \text{DL-GEP}(\gamma, \delta, \alpha, \beta, \theta, \tau)$, sa densité $f(y)$ est telle que $yf(y) \in L_{\gamma, \delta, \alpha}(\infty)$. En effet, le terme $K(\gamma, \delta, \alpha, \beta, \theta) \cdot y(2\tau(|y/\tau| + e^\theta))^{-1} \cdot (\log(\log(|y/\tau| + e^\theta)))^{-\beta}$ est une fonction à variation L-lente, pour toute valeur des paramètres. De plus, le terme e^θ n'a pas d'impact sur les ailes du terme $\exp[-\delta(\log(|y/\tau| + e^\theta))^\gamma](\log(|y/\tau| + e^\theta))^{-\alpha}$. La condition i) du théorème 1 peut alors être vérifiée facilement. En effet, en s'assurant simplement que $\gamma < 1$, l'inférence *a posteriori* sera robuste en présence de valeurs aberrantes, si ces dernières sont en nombre minoritaire.

Dans un cadre de régression linéaire, la densité du terme d'erreur est habituellement unimodale et symétrique. L'unimodalité de la densité DL-GEP est garantie, sauf si

$\alpha < 0$ et/ou $\beta < 0$. Dans ce cas, il suffit de choisir une valeur de θ assez grande afin d'obtenir l'unimodalité. Plus précisément, θ doit être tel que $\theta + \delta\theta^\gamma\gamma + \beta/\log\theta + \alpha \geq 0$. Ce critère vise à satisfaire $f'(y) \leq 0$ pour tout $y \geq 0$, puisque f est symétrique.

Définissons maintenant quelques cas particuliers intéressants.

Définition 7. Une variable aléatoire Y a une distribution DL-Pareto, notée $Y \sim \text{DL-Pareto}(\phi, \theta, \tau)$, si $Y \sim \text{DL-GEP}(0, 0, \phi + 1, 0, \theta, \tau)$, avec une densité donnée par

$$f(y) = \phi\theta^\phi(2\tau(|y/\tau| + e^\theta))^{-1}(\log(|y/\tau| + e^\theta))^{-(\phi+1)},$$

où $y \in \mathbb{R}$, $\phi > 0$, $\theta > 0$ et $\tau > 0$. La fonction de répartition de la variable aléatoire Y est donnée par

$$F(y) = \begin{cases} (1/2)\theta^\phi (\log(-y/\tau + e^\theta))^{-\phi} & \text{si } y \leq 0, \\ 1 - (1/2)\theta^\phi (\log(y/\tau + e^\theta))^{-\phi} & \text{si } y > 0. \end{cases}$$

De plus, une variable aléatoire peut être générée facilement par la méthode de simulation inverse. En effet,

$$y = \begin{cases} -\tau [\exp(\theta(2u)^{-1/\phi}) - \exp(\theta)] & \text{si } u \leq 0.5, \\ \tau [\exp(\theta(2(1-u))^{-1/\phi}) - \exp(\theta)] & \text{si } u > 0.5, \end{cases}$$

où u est la réalisation d'une variable aléatoire avec une distribution uniforme sur l'intervalle $(0, 1)$.

On a alors que $yf(y) \in L_{0,0,\phi+1}(\infty)$ et la condition i) du théorème 1 est respectée, ce qui engendre une inférence robuste en présence de valeurs aberrantes en nombre minoritaire. De plus, cette distribution est unimodale pour toute valeur des paramètres. La densité de départ pour ce cas particulier est une *Shifted-Pareto* avec paramètres ϕ et θ . Sa densité est donnée par

$$g(v) = \frac{\phi\theta^\phi}{v^{\phi+1}}, v > \theta, \phi > 0, \theta > 0,$$

soit une distribution $\text{GEP}^+(0, 0, \phi + 1, 0, \theta)$ avec $K(0, 0, \phi + 1, 0, \theta) = \phi\theta^\phi$.

Définition 8. Une variable aléatoire Y a une distribution DL-GenGamma, notée $Y \sim \text{DL-GenGamma}(\gamma, \eta, \phi, \theta, \tau)$, si $Y \sim \text{DL-GEP}(\gamma, \eta^\gamma, 1 - \phi, 0, \theta, \tau)$ avec une densité donnée par

$$f(y) = \frac{\gamma \eta^\phi}{\Gamma(\phi/\gamma, (\eta\theta)^\gamma)} \cdot (2\tau(|y/\tau| + e^\theta))^{-1} \exp[-\eta^\gamma(\log(|y/\tau| + e^\theta))^\gamma] \\ \times (\log(|y/\tau| + e^\theta))^{\phi-1},$$

où $y \in \mathbb{R}, \gamma > 0, \eta > 0, \phi > 0, \theta > 0$ ($\theta \geq 0$ si $\phi = 1$), et $\tau > 0$. Le terme $\Gamma(a, b) = \int_b^\infty u^{a-1} e^{-u} du = b^a \int_1^\infty v^{a-1} e^{-bv} dv$ où $a > 0$ et $b \geq 0$ est la fonction Gamma incomplète, avec $\Gamma(a, 0) = \Gamma(a)$, soit la fonction Gamma usuelle évaluée en a . La fonction de répartition de la variable aléatoire Y est donnée par

$$F(y) = \begin{cases} \frac{\Gamma(\phi/\gamma, (\eta \log(-y/\tau + e^\theta))^\gamma)}{2\Gamma(\phi/\gamma, (\eta\theta)^\gamma)} & \text{si } y \leq 0, \\ 1 - \frac{\Gamma(\phi/\gamma, (\eta \log(y/\tau + e^\theta))^\gamma)}{2\Gamma(\phi/\gamma, (\eta\theta)^\gamma)} & \text{si } y > 0. \end{cases}$$

On a alors que $yf(y) \in L_{\gamma, \eta^\gamma, 1-\phi}(\infty)$ et la condition i) du théorème 1 est respectée seulement lorsque $\gamma < 1$. La densité est unimodale, sauf possiblement lorsque $\phi > 1$, où il s'agit dans ce cas de choisir θ suffisamment grand, c'est-à-dire tel que $\theta + (\eta\theta)^\gamma \geq \phi - 1$. La distribution de départ pour ce cas particulier est l'aile normalisée d'une loi gamma généralisée avec une puissance γ , un paramètre de forme ϕ et un paramètre d'échelle de $1/\eta$. Sa densité est donnée par

$$g(v) = \frac{\gamma \eta^\phi}{\Gamma(\phi/\gamma, (\eta\theta)^\gamma)} \cdot v^{\phi-1} e^{-(\eta v)^\gamma}, v > \theta, \gamma > 0, \eta > 0, \phi > 0,$$

avec les mêmes contraintes que ci-haut sur θ . Il s'agit d'une distribution $\text{GEP}^+(\gamma, \eta^\gamma, 1 - \phi, 0, \theta)$, avec $K(\gamma, \eta^\gamma, 1 - \phi, 0, \theta) = \gamma \eta^\phi / \Gamma(\phi/\gamma, (\eta\theta)^\gamma)$. Puisque $\Gamma(\phi/\gamma, (\eta\theta)^\gamma) = \int_{(\eta\theta)^\gamma}^\infty u^{\phi/\gamma-1} e^{-u} du = \int_\theta^\infty \gamma \eta^\phi v^{\phi-1} e^{-(\eta v)^\gamma} dv$, il fait en sorte que la distribution soit normalisée.

Deux cas particuliers intéressants découlent de la DL-GenGamma. D'abord, il y a la distribution DL-Gamma(η, ϕ, θ, τ) lorsque $\gamma = 1$ qui provient de l'aile normalisée d'une

densité gamma. Sa densité est donnée par

$$\begin{aligned} f(y) &= \frac{\eta^\phi}{\Gamma(\phi, \eta\theta)} \cdot (2\tau(|y/\tau| + e^\theta))^{-1} \exp[-\eta \log(|y/\tau| + e^\theta)] \\ &\quad \times (\log(|y/\tau| + e^\theta))^{\phi-1} \\ &= \frac{\eta^\phi}{\Gamma(\phi, \eta\theta)} \cdot \frac{(\log(|y/\tau| + e^\theta))^{\phi-1}}{2\tau(|y/\tau| + e^\theta)^{\eta+1}}, \end{aligned}$$

avec $y \in \mathbb{R}$ et les mêmes contraintes que ci-haut pour les paramètres de la densité. La densité DL-Gamma est unimodale, sauf possiblement lorsque $\phi > 1$, où il s'agit alors de choisir une valeur pour θ telle que $\theta \geq (\phi - 1)/(\eta + 1)$. Notons que si on posait $\theta = 0$ (ce que les contraintes excluent sauf si $\phi = 1$), nous obtiendrions une densité correspondant à une loi log-gamma qui aurait été déplacée vers l'origine puis doublée sur les réels. Elle serait toutefois bimodale avec $f(0) = 0$ si $\phi > 1$ ou discontinue en 0 avec $\lim_{y \rightarrow 0} f(y) = \infty$ si $\phi < 1$.

Ensuite, il y a la distribution DL-exponentielle(η, θ, τ) lorsque $\gamma = \phi = 1$ qui provient de l'aile normalisée d'une densité exponentielle. Sa densité est donnée par

$$f(y) = \frac{\eta}{e^{-\eta\theta}} \cdot \frac{1}{2\tau(|y/\tau| + e^\theta)^{\eta+1}},$$

avec $y \in \mathbb{R}$ et les mêmes contraintes que ci-haut pour les paramètres de la densité. Cette densité est toujours unimodale.

Définition 9. Une variable aléatoire Y a une distribution DL-normale, notée $Y \sim DL\text{-normale}(\phi, \theta, \tau)$, si $Y \sim DL\text{-GEP}(2, 1/(2\phi^2), 0, 0, \theta, \tau)$ avec une densité donnée par

$$\begin{aligned} f(y) &= ((1 - \Phi(\theta/\phi))\sqrt{2\pi}\phi)^{-1} (2\tau(|y/\tau| + e^\theta))^{-1} \\ &\quad \times \exp \left[-\frac{1}{2\phi^2} \left(\log(|y/\tau| + e^\theta) \right)^2 \right], \end{aligned}$$

où $y \in \mathbb{R}, \phi > 0, \theta \geq 0, \tau > 0$ et $\Phi(a) = \int_{-\infty}^a (1/\sqrt{2\pi})e^{-u^2/2} du$, soit la fonction de répartition d'une loi normale centrée réduite. La fonction de répartition de la variable aléatoire Y est donnée par

$$F(y) = \begin{cases} \frac{1 - \Phi(\log(-y/\tau + e^\theta)/\phi)}{2(1 - \Phi(\theta/\phi))} & \text{si } y \leq 0, \\ 1 - \frac{1 - \Phi(\log(y/\tau + e^\theta)/\phi)}{2(1 - \Phi(\theta/\phi))} & \text{si } y > 0. \end{cases}$$

On a alors que $y f(y) \in L_{2,1/(2\phi^2),0}(\infty)$ et la condition i) du théorème 1 n'est pas respectée. Cette densité est unimodale $\forall \theta \geq 0$. La distribution de départ pour ce cas particulier est l'aile normalisée d'une loi normale centrée avec un paramètre d'échelle de ϕ . Sa densité est donnée par

$$g(v) = ((1 - \Phi(\theta/\phi))\sqrt{2\pi}\phi)^{-1} \exp \left[-(1/2) (v/\phi)^2 \right], v > \theta, \phi > 0,$$

avec la même contrainte que ci-haut sur le paramètre θ . Il s'agit d'une distribution $\text{GEP}^+(2, 1/(2\phi^2), 0, 0, \theta)$ avec $K(2, 1/(2\phi^2), 0, 0, \theta) = [(1 - \Phi(\theta/\phi))\sqrt{2\pi}\phi]^{-1}$. Notons que si on posait $\theta = -\infty$ (ce que les contraintes excluent), la densité DL-normale correspondrait à une densité log-normale doublée sur les réels. Elle serait toutefois bimodale avec $f(0) = 0$. La contrainte $\theta \geq 0$ fait en sorte que la densité DL-normale est positive et unimodale.

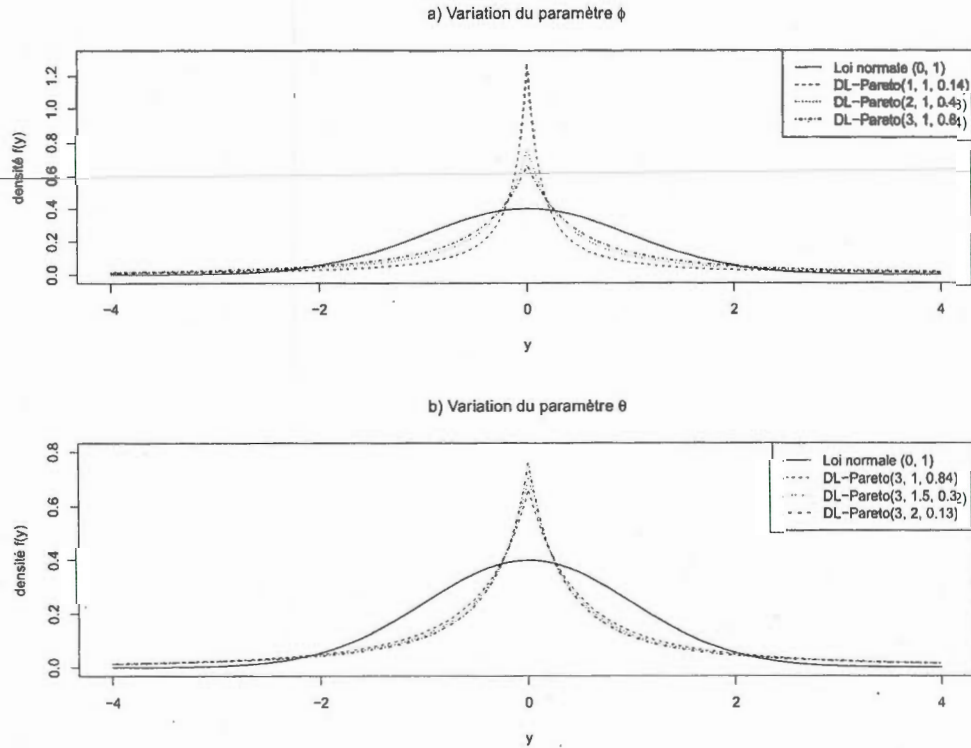


Figure 3.1 Comportement d'une densité DL-Pareto(ϕ, θ, τ) lorsqu'un seul paramètre varie à la fois.

Les figures 3.1, 3.2, 3.3 et 3.4 illustrent le comportement des différents cas particuliers de la loi DL-GEP. L'objectif est de les comparer avec une loi normale centrée réduite et de montrer l'impact des différents paramètres sur la forme des densités. Pour chaque densité, la valeur du paramètre d'échelle τ a été fixée afin que la distance interquartile soit la même que celle d'une loi normale centrée réduite. Ceci facilite la comparaison entre les différentes densités.

La figure 3.1 illustre l'impact marginal de la variation d'un paramètre de la distribution DL-Pareto. Le paramètre ϕ influence l'épaisseur des ailes et la hauteur du mode. Plus la valeur de ϕ est petite, plus les ailes sont relevées et le mode est haut. Le paramètre θ , quant à lui, n'a pas d'effet considérable sur la forme de la densité, ce qui indique que son impact est similaire au paramètre d'échelle τ . En effet, les comparaisons sont effectuées suite à l'ajustement de τ faisant en sorte que toutes les densités aient des distances interquartiles identiques.

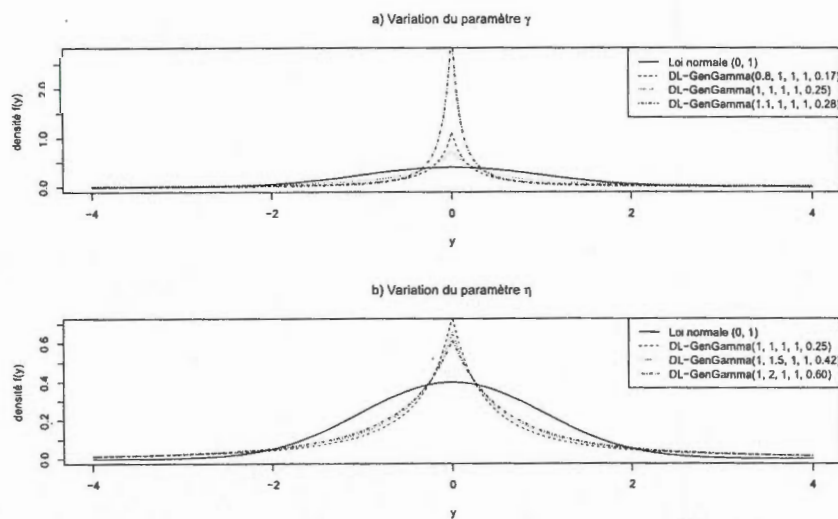


Figure 3.2 Comportement d'une densité DL-GenGamma($\gamma, \eta, \phi, \theta, \tau$) lorsqu'un seul paramètre varie à la fois (partie 1).

Les figures 3.2 et 3.3 illustrent l'impact marginal de la variation d'un paramètre de la distribution DL-G-Gamma. Les trois premiers paramètres, soient γ, η et ϕ , influencent

l'épaisseur des ailes. Le paramètre γ a plus d'impact que les deux autres. Rappelons que la distribution respecte la condition i) du théorème 1 si $\gamma < 1$, ce qui engendre une inférence robuste en présence de valeurs aberrantes en nombre minoritaire. De plus, on remarque que le paramètre γ influence considérablement la hauteur du mode et la forme de la densité, ce qui est beaucoup moins marqué pour les paramètres η et ϕ . Le paramètre θ , quant à lui, peut influencer la forme de la densité près de 0, lorsque sa valeur fait en sorte que la densité soit unimodale ou non. La densité $\text{DL-GenGamma}(1, 1, 2, 0.5, \tau)$ est un exemple de cas limite entre l'unimodalité et la bimodalité, puisqu'elle satisfait $\theta = (\phi - 1)/(\eta + 1)$. Lorsque $\theta + (\eta\theta)^\gamma \geq \phi - 1$, on a l'unimodalité et le paramètre θ n'a plus d'impact marqué sur la forme de la densité.

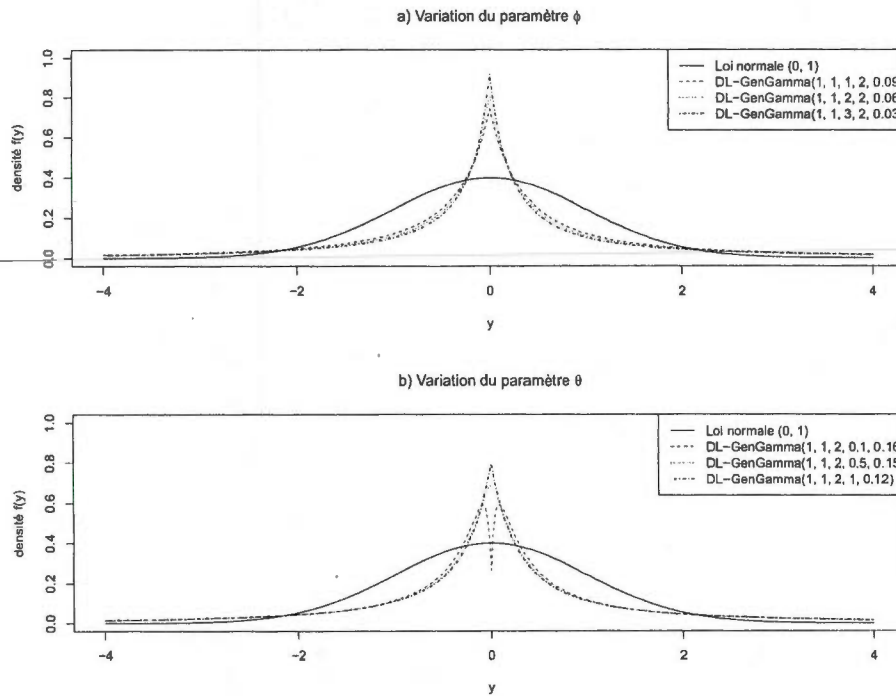


Figure 3.3 Comportement d'une densité $\text{DL-GenGamma}(\gamma, \eta, \phi, \theta, \tau)$ lorsqu'un seul paramètre varie à la fois (partie 2).

On peut remarquer aux figures 3.2 et 3.3 que si $\gamma = 1$, la distribution DL-GenGamma

devient une loi DL-Gamma et si $\gamma = \phi = 1$, la distribution DL-GenGamma devient une loi DL-exponentielle.

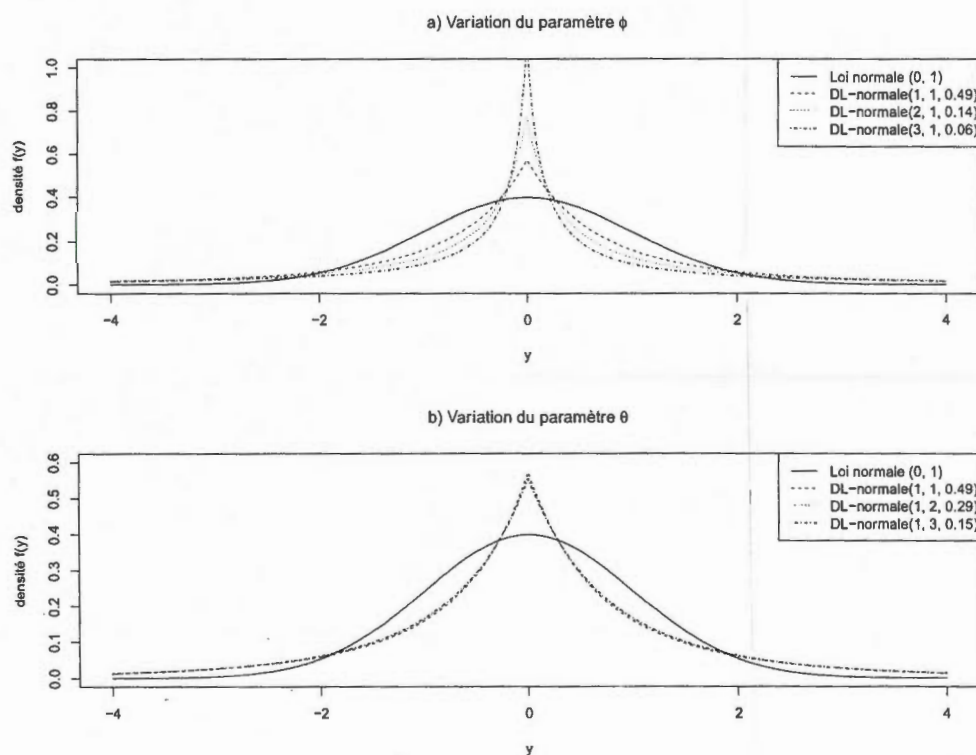


Figure 3.4 Comportement d'une densité DL-normale(ϕ, θ, τ) lorsqu'un seul paramètre varie à la fois.

Pour la figure 3.4, des commentaires semblables à ceux de la distribution DL-Pareto peuvent être effectués. Le paramètre ϕ influence l'épaisseur des ailes, plus sa valeur est petite, plus les ailes sont relevées. Le paramètre θ , quant à lui, n'a pas d'effet considérable sur la forme de la densité, ce qui indique que son impact est similaire au paramètre d'échelle τ .

Finalement, la figure 3.5 illustre une des densités propres ayant les ailes les plus relevées, soit la distribution DL-GEP(0, 0, 1, 2, 2, 0.03). On remarque son mode très haut et ses ailes rapprochées très basses.

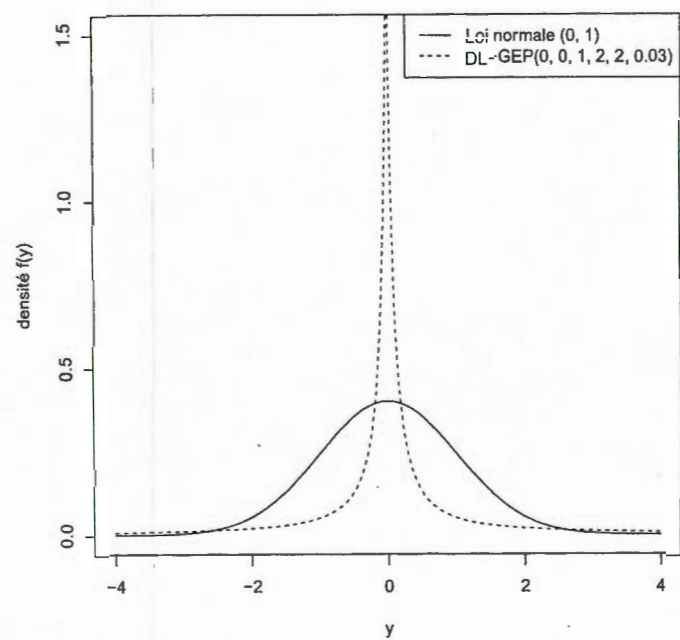


Figure 3.5 Comportement d'une densité DL-GEP(0, 0, 1, 2, 2, 0.03).

CHAPITRE IV

EXEMPLE À PARTIR DE DONNÉES SIMULÉES

Au chapitre 2, nous avons montré que l'inférence *a posteriori* est robuste face à des valeurs aberrantes en nombre minoritaire, sous certaines conditions. Dans le présent chapitre, les résultats sont illustrés à l'aide de données simulées. Un cas non robuste est tout d'abord présenté afin d'illustrer l'impact que peuvent avoir des observations non représentatives sur l'inférence *a posteriori*. L'hypothèse habituelle de normalité des erreurs est alors utilisée. Ensuite, le cas robuste est présenté, où le terme d'erreur a une distribution DL-Pareto(ϕ, θ, τ) puisque cette densité a des ailes suffisamment relevées, satisfaisant ainsi la condition i) du théorème 1.

Les paramètres β et σ du modèle de régression sont estimés ponctuellement par la médiane *a posteriori*. L'espérance n'est pas calculée puisque qu'elle n'existe pas dans le cas où les erreurs ont une distribution DL-Pareto. L'intervalle de crédibilité à plus haute densité *a posteriori* (IC PHDP) à 95% est aussi calculé. Ces estimations sont calculées à partir des densités *a posteriori* marginales de β et σ .

4.1 Présentation des données

Le modèle de régression utilisé pour simuler les données sans valeurs aberrantes est $Y_i = \beta X_i + \epsilon_i, i = 1, \dots, n$. La taille de l'échantillon a été fixée à $n = 10$. Les observations de la variable explicative ont été générées à partir d'une loi uniforme sur l'intervalle $(0, 10)$, puis les observations de la variable d'intérêt ont été générées avec $\beta = 1$ et une

erreur distribuée selon une loi normale centrée réduite. Les paramètres du modèle ont été choisis arbitrairement, tout en s'assurant que la variabilité du terme d'erreur ne soit pas trop élevée par rapport au paramètre β . Ainsi, la tendance linéaire est claire.

i	1	2	3	4	5	6	7	8	9	10
x_i	0.3	1.7	2.4	3.5	4.9	5.9	6.7	7.4	8.3	9.9
y_i	1.2	1.6	2.2	4.9	5.4	4.3	6.7	6.3	8.1	10.4
z_i	4.00	0.94	0.92	1.40	1.10	0.73	1.00	0.85	0.98	1.05

Tableau 4.1 Échantillon sans données aberrantes.

i	1	2	3	4	5	6	7	8	9	10
ϵ_i	0.9	-0.1	-0.2	1.4	0.5	-1.6	0.0	-1.1	-0.2	-0.5
ϵ_i/x_i	3.00	-0.06	-0.08	0.40	0.10	-0.27	0.00	-0.15	-0.02	0.05
$1/ x_i $	3.33	0.59	0.42	0.29	0.20	0.17	0.15	0.14	0.12	0.10

Tableau 4.2 Mesures expliquant la variabilité des observations z dans l'échantillon sans données aberrantes.

Les données générées, x_i, y_i et $z_i = y_i/x_i$ sont présentées dans le tableau 4.1. L'erreur $\epsilon_i = y_i - \beta x_i$, représentant la distance verticale entre le point (x_i, y_i) et la droite $Y = \beta X$, et l'erreur $\epsilon_i/x_i = z_i - \beta$, représentant la distance verticale entre le point (x_i, z_i) et la droite $Z = \beta$, sont présentées dans le tableau 4.2. Le paramètre d'échelle $1/|x_i|$ est aussi présenté, il est l'écart-type de ϵ_i/x_i , puisque ϵ_i a une distribution normale centrée réduite. Il permet de mieux apprécier la nature aberrante d'un point (x_i, z_i) . Rappelons que β a été fixé à 1.

La figure 4.1 a) illustre l'échantillon sous la forme des (x_i, y_i) avec la droite $Y = X$ et la figure 4.1 b) illustre l'échantillon sous la forme des (x_i, z_i) avec la droite $Z = 1$. Le vecteur z semble, à première vue, contenir une valeur extrême, soit $z_1 = 4.00$ avec $\epsilon_1/x_1 = 3.00$. Cependant, tel qu'indiqué dans le tableau 4.2, ϵ_1/x_1 a été générée d'une distribution normale centrée avec un écart-type de 3.33. Une valeur de $\epsilon_1/x_1 = 3.00$ est alors plausible. C'est la combinaison de l'erreur avec une petite valeur de $|x_1|$ qui crée

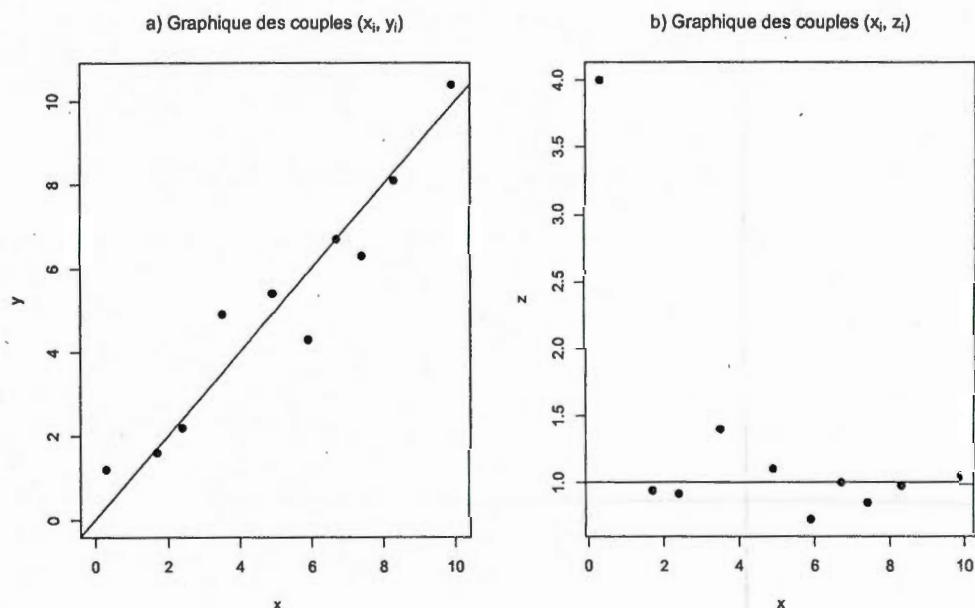


Figure 4.1 Échantillon sans données aberrantes.

l'effet graphique « extrême ».

L'analyse est d'abord effectuée sur les données d'origine, sans valeurs extrêmes, puis des observations sont modifiées afin de les rendre clairement aberrantes. Successivement, une observation $z_i = y_i/x_i$ est déplacée vers une valeur de 100, soit une valeur très élevée comparativement au reste de l'échantillon. En effet, les valeurs de $1/|x_i|$ varient entre 0.10 et 3.33 et celles des z_i entre 0.73 et 4.00.

Les densités *a posteriori* des paramètres sont estimées à partir d'un algorithme de Gibbs. Ainsi, à chaque itération, des intégrales numériques à une dimension doivent être calculées afin d'obtenir des réalisations de la forme $\sigma^{(t)} \mid \beta^{(t-1)}, \mathbf{z}, \mathbf{x} \sim \pi_1(\sigma \mid \beta^{(t-1)}, \mathbf{z}, \mathbf{x})$ et $\beta^{(t)} \mid \sigma^{(t)}, \mathbf{z}, \mathbf{x} \sim \pi_2(\beta \mid \sigma^{(t)}, \mathbf{z}, \mathbf{x})$. Une valeur initiale de $\beta^{(0)}$ doit être spécifiée afin de commencer l'algorithme. La médiane de $\mathbf{z} = (z_1, \dots, z_n) = (y_1/x_1, \dots, y_n/x_n)$ est utilisée puisqu'il s'agit d'une mesure robuste face aux valeurs aberrantes. L'analyse statistique est effectuée à l'aide du logiciel R Development Core Team (2011) et l'algorithme

est construit sous C++.

4.2 Cas non robuste - $\epsilon_i \sim \mathcal{N}(0, 1)$

Dans cette section, l'hypothèse habituelle de normalité des erreurs est posée. Les erreurs $\epsilon_i \mid \sigma$ sont conditionnellement indépendantes et identiquement distribuées avec une densité donnée par

$$(1/\sigma)f(\epsilon_i/\sigma) = (\sqrt{2\pi}\sigma)^{-1} \exp \left[-(1/2) (\epsilon_i/\sigma)^2 \right], i = 1, \dots, 10,$$

où $\epsilon_i \in \mathbb{R}$ et $\sigma > 0$. Dans ce cas, $\epsilon_i f(\epsilon_i) \notin L_{\gamma, \delta, \alpha}(\infty)$, puisque ses ailes sont moins relevées que toutes fonctions appartenant à $L_{\gamma, \delta, \alpha}(\infty)$. La condition i) du théorème 1 n'est donc pas satisfaite sous cette hypothèse. Alors, il n'est pas garanti que l'inférence *a posteriori* sera robuste face à des valeurs aberrantes.

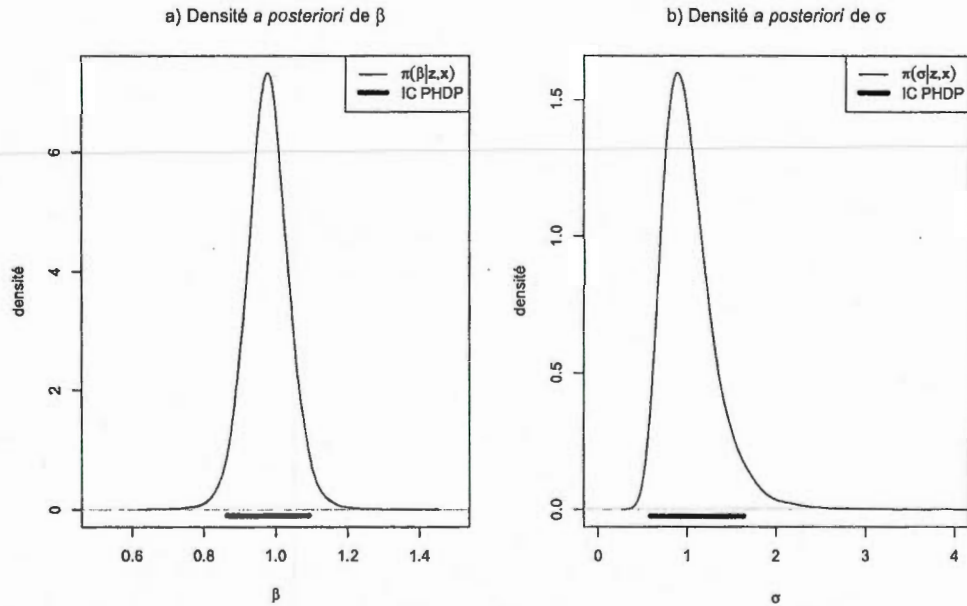


Figure 4.2 Densités *a posteriori* avec les données sans valeurs aberrantes et sous l'hypothèse de normalité.

En se basant sur l'échantillon sans données aberrantes et sur l'hypothèse de normalité, l'estimation de β par la médiane *a posteriori* est égale à 0.98 et l'IC PHDP est donné

par (0.87, 1.09). Pour ce qui est de σ , l'estimation par la médiane *a posteriori* est égale à 0.98, et l'IC PHDP est donné par (0.59, 1.64). Ces mesures sont en accord avec le modèle générateur ($\beta = \sigma = 1$), ce qui était attendu puisqu'il n'y a pas de valeurs aberrantes. Par ailleurs, la figure 4.2 illustre la symétrie de la densité de β et l'asymétrie de celle de σ avec une aile de droite plus relevée.

Maintenant, un échantillon contaminé par une valeur aberrante est considéré. Il est constitué des neuf premiers couples d'origine et du couple $(x_{10}, y_{10}) = (9.9, 990)$ ou $(x_{10}, z_{10}) = (9.9, 100)$.

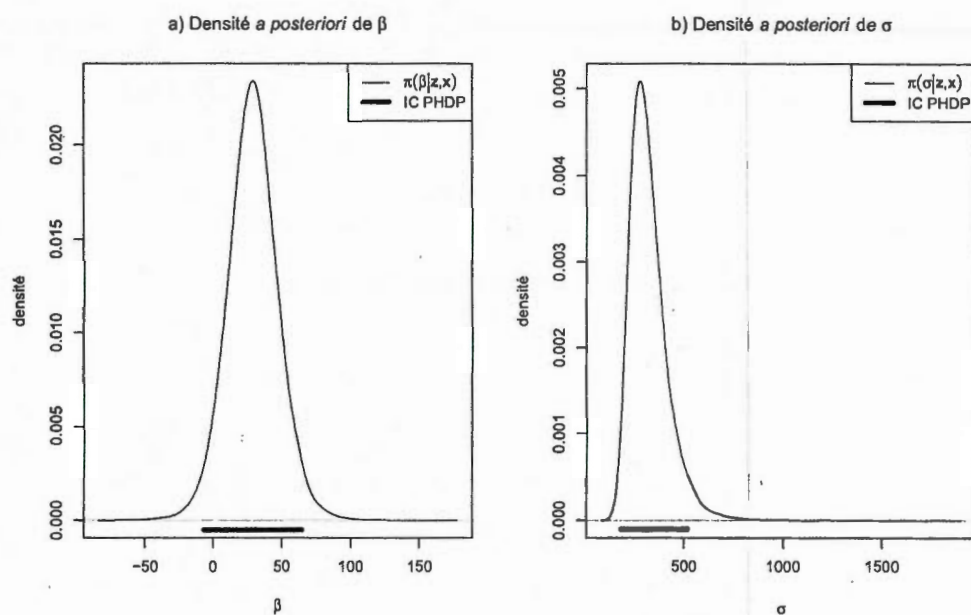


Figure 4.3 Densités *a posteriori* avec les couples 1 à 9 d'origine et $(x_{10}, y_{10}) = (9.9, 990)$, sous l'hypothèse de normalité.

La figure 4.3 illustre les densités *a posteriori* de β et σ en présence de l'échantillon contaminé par une valeur aberrante. L'estimation de β par la médiane *a posteriori* est égale à 29 et l'IC PHDP est donné par (-6.62, 64.69). On remarque que la densité de β est à peu près symétrique. Pour ce qui est de σ , l'estimation par la médiane *a posteriori* est égale à 307, et l'IC PHDP est donné par (185.76, 516.37). La présence

d'une valeur extrême a donc un sérieux impact sur l'inférence *a posteriori*. En effet, les mesures calculées sont en désaccord avec la majorité des observations. En particulier pour l'estimation de β , les neuf premiers couples ont une valeur z_i variant entre 0.73 et 4.00, et $z_{10} = 100$. Aucune d'entre elles ne se trouvent dans les environs de l'estimation ponctuelle de β qui est égale à 29.

Dans la prochaine section, les résultats de l'inférence engendrée par une distribution ayant des ailes suffisamment relevées sont présentés. Cette caractéristique des ailes limite l'influence des valeurs aberrantes.

4.3 Cas robuste - $\epsilon_i \sim \text{DL-Pareto}(\phi, \theta, \tau)$

Dans cette section, on suppose que les erreurs ont une distribution $\text{DL-Pareto}(\phi, \theta, \tau)$ avec une densité donnée par

$$f(\epsilon_i) = \phi\theta^\phi (2\tau(|\epsilon_i/\tau| + e^\theta))^{-1} (\log(|\epsilon_i/\tau| + e^\theta))^{-(\phi+1)},$$

où $\epsilon_i \in \mathbb{R}$, $\phi > 0$, $\theta > 0$ et $\tau > 0$. Cette distribution respecte la condition i) du théorème 1 avec une densité unimodale pour toute valeur des paramètres. Le paramètre ϕ est fixé arbitrairement à 3. Sa valeur affecte l'épaisseur des ailes et la hauteur du mode, comme le suggère la figure 3.1. Une étude approfondie permettrait de connaître son impact sur l'inférence. Le paramètre θ est fixé arbitrairement à 1. Comme mentionné au chapitre 3, sa valeur a environ le même impact sur la densité que celui de τ .

Afin de faciliter la comparaison avec les résultats du cas non robuste de la section 4.2, le paramètre τ de la distribution DL-Pareto est choisi de façon à ce que la distance interquartile soit la même que celle d'une loi normale centrée réduite, soit $\tau = 0.835943$. La distribution $\text{DL-Pareto}(\phi = 3, \theta = 1, \tau = 0.835943)$ peut être comparée graphiquement à la loi normale centrée réduite en se référant à la figure 3.1.

En se basant sur l'échantillon de départ, l'estimation de β par la médiane *a posteriori* est égale à 0.99, et l'IC PHDP est donné par (0.90, 1.07). Ces valeurs sont très près de ce qui a été trouvé en supposant la normalité des erreurs. Pour ce qui est de σ , l'estima-

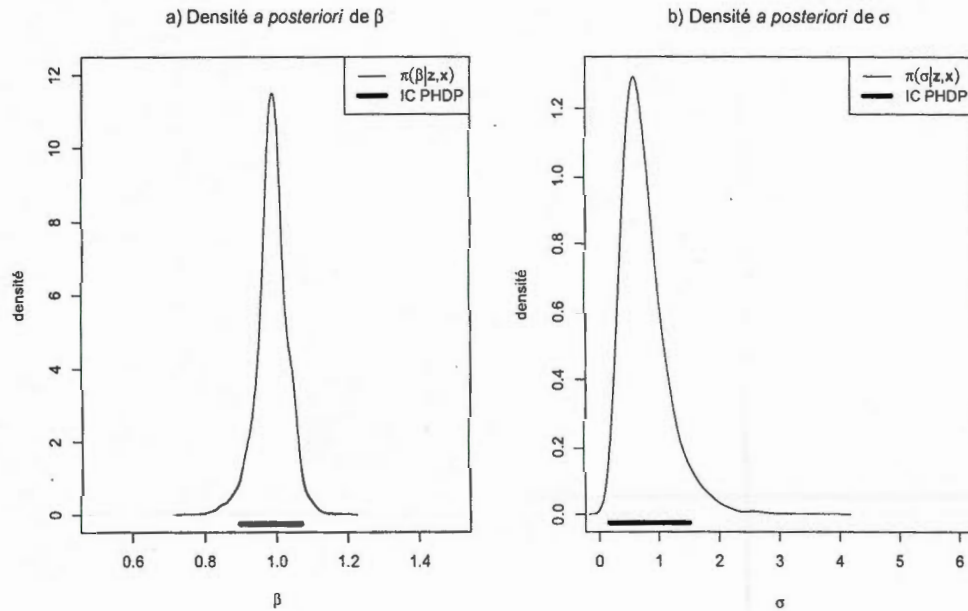


Figure 4.4 Densités *a posteriori* avec les données sans valeurs aberrantes.

tion par la médiane *a posteriori* est égale à 0.68. L'IC PHDP est donné par (0.17, 1.51). Comparativement au cas non robuste, ces valeurs sont quelques peu différentes. Il n'y a cependant pas de contradiction puisque la signification du paramètre σ n'est pas la même dans les deux cas. En s'assurant que la distance interquartile de la distribution DL-Pareto est la même que celle d'une loi normale centrée réduite, on s'attendait uniquement à ce que les estimations de σ soient dans un même ordre de grandeur. On remarque d'ailleurs que les densités *a posteriori* de σ sont semblables dans les deux cas. Par ailleurs, la figure 4.4 illustre l'asymétrie de la densité de σ avec une aile de droite plus relevée. On remarque aussi que la densité de β est à peu près symétrique.

L'analyse sur l'échantillon contaminé par une valeur aberrante est maintenant présentée. L'échantillon est constitué des neuf premiers couples d'origine et du couple $(x_{10}, y_{10}) = (9.9, 990)$ ou $(x_{10}, z_{10}) = (9.9, 100)$. L'estimation de β par la médiane *a posteriori* est égale à 0.97, soit la même valeur (arrondie à deux décimales) que celle basée sur la densité excluant la valeur extrême. Les IC PHDP pour les échantillons incluant et

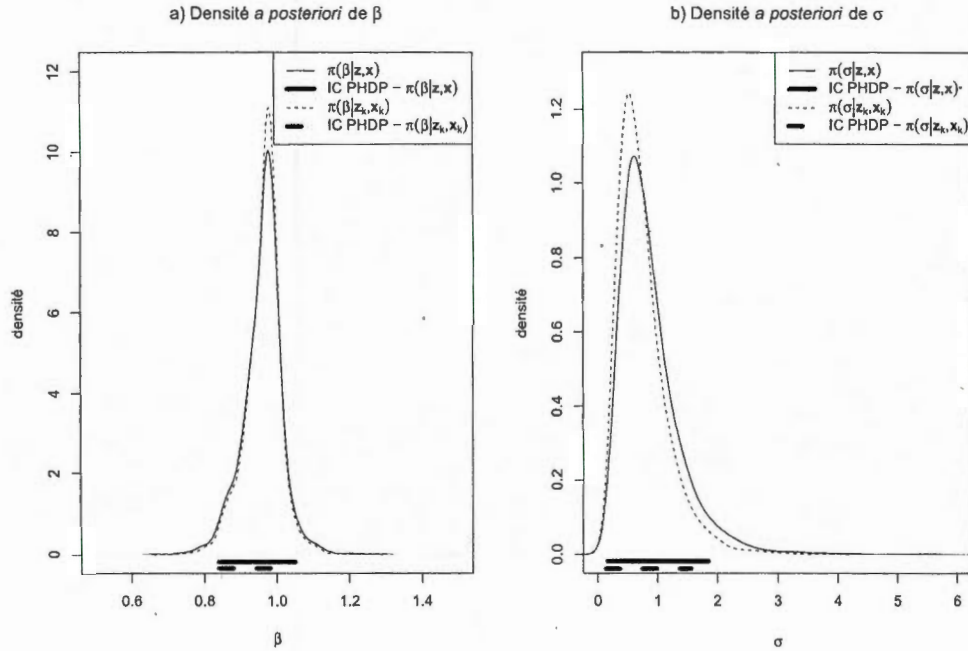


Figure 4.5 Densités *a posteriori* avec les couples 1 à 9 d'origine et $(x_{10}, y_{10}) = (9.9, 990)$.

excluant la valeur aberrante sont donnés respectivement par $(0.84, 1.05)$ et $(0.84, 1.04)$. Ceci illustre bien la convergence des deux densités lorsqu'une valeur aberrante s'éloigne vers l'infini. Pour ce qui est de σ , l'estimation par la médiane *a posteriori* est égale à 0.76. L'IC PHDP est donné par $(0.17, 1.85)$. L'estimation par la médiane basée sur l'échantillon excluant la valeur extrême est 0.66 et l'IC PHDP est $(0.14, 1.56)$. Ceci illustre encore la robustesse, mais aussi la convergence plus lente pour le paramètre d'échelle comparativement à celle du paramètre de position. Par ailleurs, la figure 4.5 illustre l'asymétrie de la densité de σ avec une aile de droite plus relevée. On remarque aussi que la densité de β est à peu près symétrique.

L'analyse sur l'échantillon contaminé par deux valeurs aberrantes est maintenant présentée. L'échantillon est constitué des huit premiers couples d'origine et des couples $(x_9, y_9) = (8.3, 830)$ ou $(x_9, z_9) = (8.3, 100)$ et $(x_{10}, y_{10}) = (9.9, 990)$ ou $(x_{10}, z_{10}) = (9.9, 100)$. L'estimation de β par la médiane *a posteriori* est égale à 0.94, soit la même

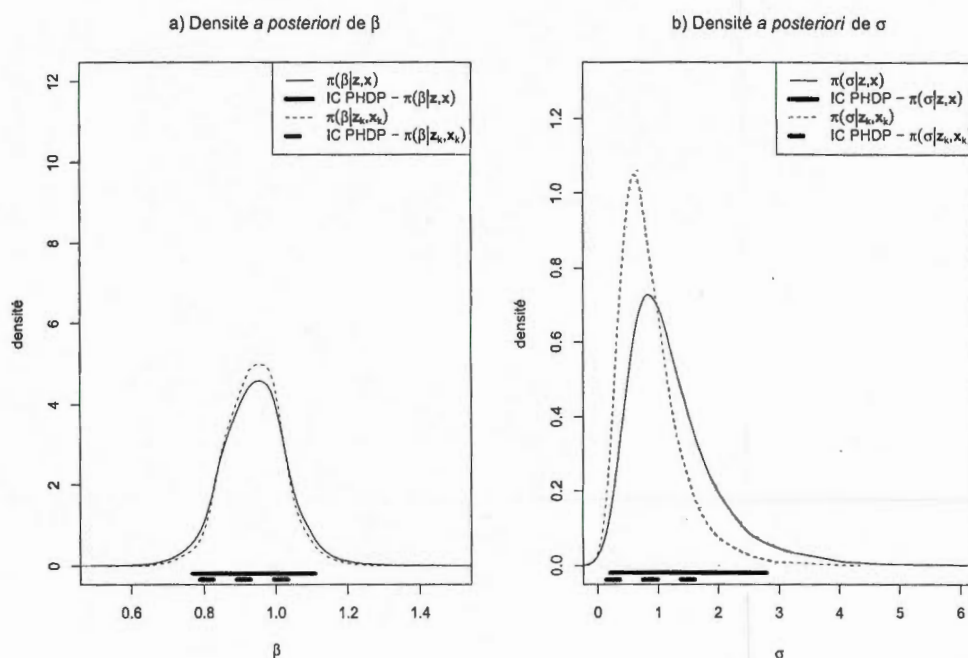


Figure 4.6 Densités *a posteriori* avec les couples 1 à 8 d'origine, $(x_9, y_9) = (8.3, 830)$ et $(x_{10}, y_{10}) = (9.9, 990)$.

valeur (arrondie à deux décimales) que celle basée sur la densité qui exclut la valeur extrême. Les IC PHDP pour les échantillons incluant et excluant la valeur aberrante sont donnés respectivement par $(0.77, 1.11)$ et $(0.79, 1.09)$. Pour ce qui est de σ , l'estimation par la médiane *a posteriori* est égale à 1.08. L'IC PHDP est donné par $(0.21, 2.79)$. L'estimation par la médiane basée sur l'échantillon excluant la valeur extrême est 0.77 et l'IC PHDP est $(0.15, 1.85)$. On remarque encore une fois la convergence des densités pour le paramètre de position et le paramètre d'échelle. Elle est plus lente pour le paramètre d'échelle et une donnée aberrante supplémentaire ralentit davantage cette convergence. Par ailleurs, la figure 4.6 illustre l'asymétrie de la densité de σ avec une aile de droite plus relevée. On remarque aussi que la densité de β est à peu près symétrique.

La convergence des densités $\pi(\cdot | \mathbf{z}, \mathbf{x})$ vers $\pi(\cdot | \mathbf{z}_k, \mathbf{x}_k)$ lorsque les valeurs aberrantes tendent vers l'infini est représentée à la figure 4.7 par la convergence de quelques mesures

telles la médiane et les bornes de l'intervalle de crédibilité. Les carrés aux extrémités de cette figure représentent les estimations basées sur $\pi(\beta|z_k, \mathbf{x}_k)$ et $\pi(\sigma|z_k, \mathbf{x}_k)$. Comme mentionné antérieurement, la convergence est plus lente pour le paramètre d'échelle que pour le paramètre de position.

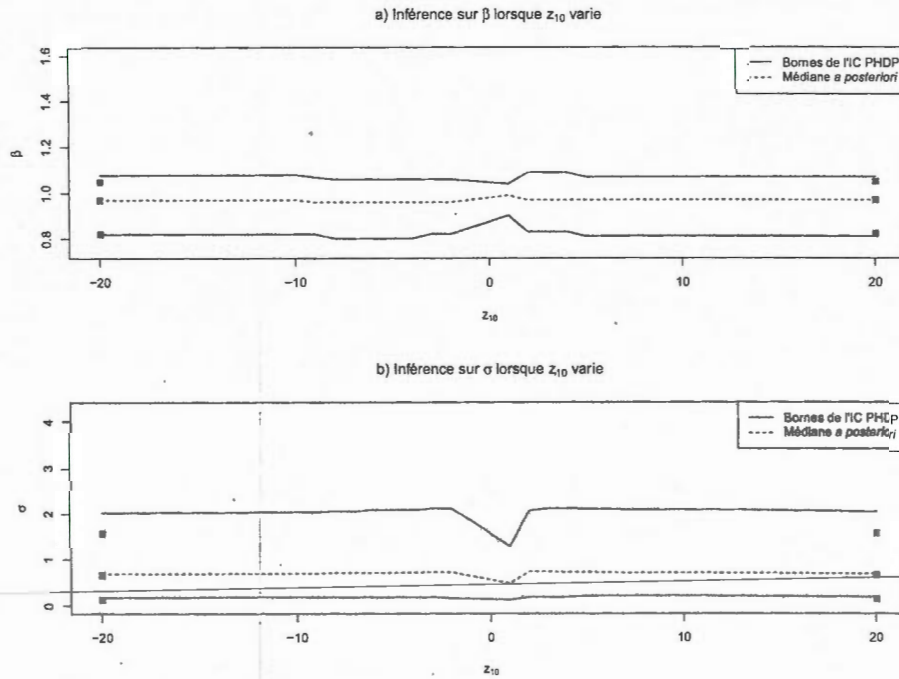


Figure 4.7 Illustration de la convergence de mesures basées sur l'échantillon complet (avec une donnée aberrante) vers celles excluant la valeur aberrante.

La figure 4.8 illustre l'impact du choix de distribution sur l'inférence de β en présence d'une valeur aberrante. Sous l'hypothèse de normalité, une valeur aberrante a un impact considérable sur l'estimation ponctuelle de β . Dans le cas de la distribution DL-Pareto, la valeur aberrante est rapidement écartée.

Dans ce chapitre, l'impact de la distribution du terme d'erreur sur l'inférence *a posteriori* a été étudié. En effet, si on suppose que le terme d'erreur a une distribution ayant des ailes suffisamment relevées, par exemple la distribution DL-Pareto, l'inférence est robuste face à des valeurs extrêmes en nombre minoritaire. De plus, cette distribution

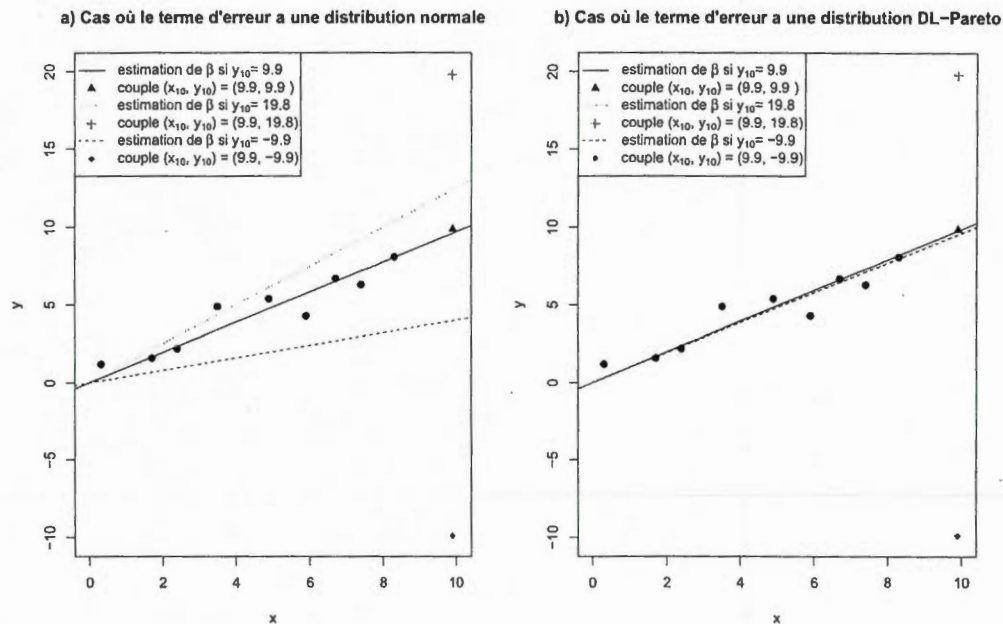


Figure 4.8 Illustration de l'impact du choix de distribution sur l'inférence en présence d'une valeur aberrante.

procure des résultats cohérents en présence d'un échantillon sans données aberrantes. Ce n'est pas le cas lorsque la densité du terme d'erreur a des ailes ne respectant pas la condition i) du théorème 1, par exemple la loi normale. En effet, l'inférence est fortement contaminée par la présence de valeurs aberrantes. D'autre part, la vitesse de convergence de la densité *a posteriori* vers celle excluant les valeurs aberrantes, lorsque celles-ci tendent vers plus ou moins l'infini, n'est pas la même pour le paramètre de position que pour le paramètre d'échelle. Elle est plus lente dans le cas du paramètre d'échelle.

CHAPITRE V

ÉTUDE DE CAS

La gestion de risques est certainement un des sujets centraux dans le domaine financier. La diversification du portefeuille boursier, par exemple, est une procédure efficace de réduction de risque. Ainsi, la connaissance des interactions entre les actifs financiers devient cruciale.

Dans ce contexte, l'application du modèle de régression robuste utilisé dans ce mémoire est étudié. Il servira à l'analyse de l'interaction, sous un angle simplifié, des économies canadienne et américaine. Ainsi, le lien entre le rendement quotidien de l'indice S&P/TSX (un indice représentatif de l'économie canadienne) et celui du S&P 500 (un indice représentatif de l'économie américaine) est considéré, pour un mois donné. L'analyse s'appuie sur la base de données de Yahoo! Finance. Une analyse descriptive des rendements quotidiens de chacun des indices pour les dix dernières années est présentée au tableau 5.1.

	Min.	25ème perc.	Méd.	Moy.	75ème perc.	Max.	Écart-type
S&P/TSX	-9.04	-0.60	0.07	0.01	0.62	10.99	1.37
S&P 500	-9.32	-0.49	0.06	0.03	0.64	9.82	1.20

Tableau 5.1 Analyse descriptive des rendements quotidiens (%) du S&P/TSX et S&P 500 pour les dix dernières années.

On sait que historiquement il y a une corrélation positive entre les économies canadienne

et américaine. Une analyse préliminaire a été effectuée à l'aide du modèle de régression linéaire simple avec une constante non-nulle. Le modèle a été estimé par la méthode des moindres carrés où le rendement quotidien du S&P 500 est la variable explicative et celui du S&P/TSX est la variable dépendante. Quelques faits intéressants en ressortent.

- Il y a plus de 88% des mois où le lien entre les rendements quotidiens du S&P 500 et ceux du S&P/TSX est significativement positif (valeur-p inférieure à 0.05)
- Il y a moins de 3% des mois où la constante est significativement différente de 0 (valeur-p inférieure à 0.05).
- Pour 5% des mois, il y a des évidences significatives que les erreurs ne proviennent pas d'une loi normale (test de Jarque-Bera avec une valeur-p inférieure à 0.05).

Afin d'utiliser le modèle de régression linéaire simple à constante nulle, on fait l'hypothèse qu'un rendement nul du S&P 500 est associé en moyenne à un rendement nul du S&P/TSX, ce qui paraît raisonnable d'autant plus que la moyenne des rendements quotidiens de chacun des indices est très près de 0.

L'objectif de cette étude de cas est de comparer, à l'aide de données réelles, les résultats obtenus lorsque l'hypothèse classique de normalité des erreurs est posée à ceux découlant d'un choix de distribution respectant la condition i) du théorème 1. Sous cette condition, l'inférence *a posteriori* est robuste face à des valeurs aberrantes en nombre minoritaire. La distribution utilisée pour arriver à cette fin est la DL-Pareto. Le choix des paramètres est le même que celui fait dans la section 4.3. Les deux distributions peuvent être comparées graphiquement en se référant à la figure 3.1.

Notons que les intervalles de crédibilité à plus haute densité *a posteriori* (IC PHDP) calculés dans ce chapitre ont une probabilité de 95%. Tout d'abord, l'analyse d'un mois représentatif est présenté.

Les figures 5.1 et 5.2 illustrent le comportement de l'inférence *a posteriori* en présence de données qui semblent normales (valeur-p supérieure à 0.15 avec le test de Jarque-Bera). Dans la figure 5.2, l'indice de la densité π indique la nature de la distribution des erreurs

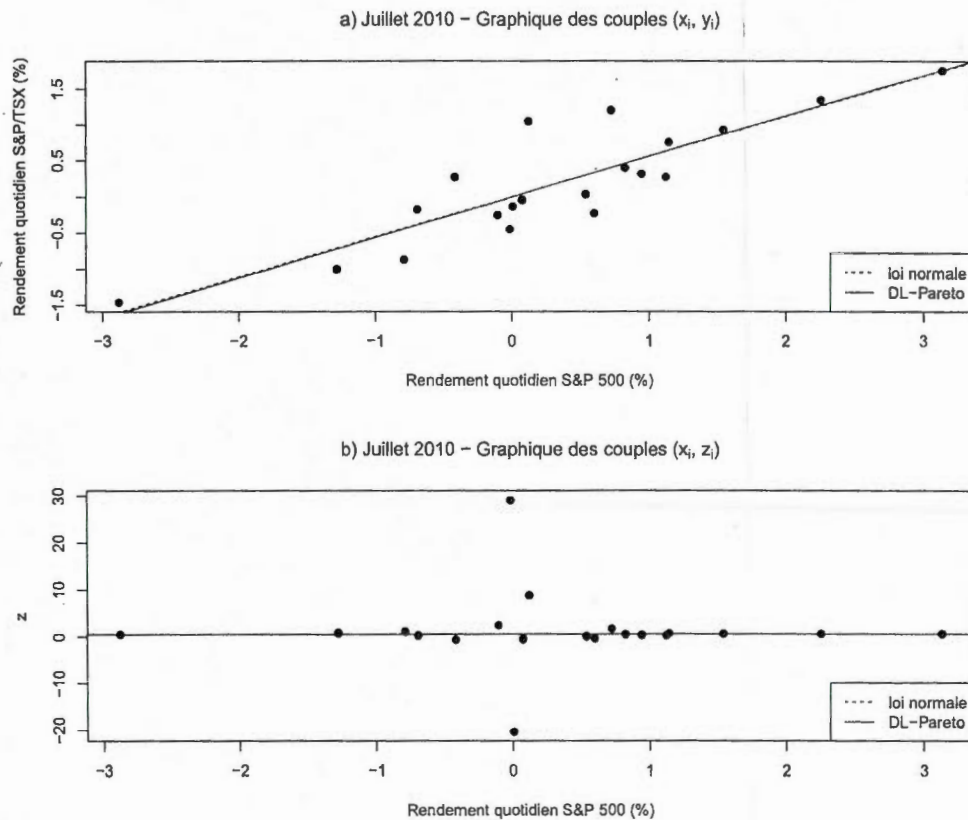


Figure 5.1 Illustration du modèle de régression où les droites sont estimées par la médiane *a posteriori*.

(normale ou DL-Pareto). L'estimation du paramètre β par la médiane *a posteriori* est la même dans les deux cas (arrondie à deux décimales), soit 0.56. La différence se situe surtout au niveau des densités *a posteriori* des paramètres. Quelques points sont légèrement éloignés de la droite, comparé à la majorité. Ceci engendre une différence au niveau de la variabilité du paramètre β , les intervalles de crédibilité sont respectivement (0.41, 0.70) et (0.45, 0.63) pour la loi normale et la distribution DL-Pareto. Pour ce qui est du paramètre σ , sous l'hypothèse de normalité, l'estimation par la médiane et l'IC PHDP sont respectivement 0.42 et (0.29, 0.58). Lorsque les erreurs sont distribuées selon une DL-Pareto, l'estimation par la médiane et l'IC PHDP sont respectivement 0.31 et (0.14, 0.54). En s'assurant que la distance interquartile de la distribution DL-Pareto est

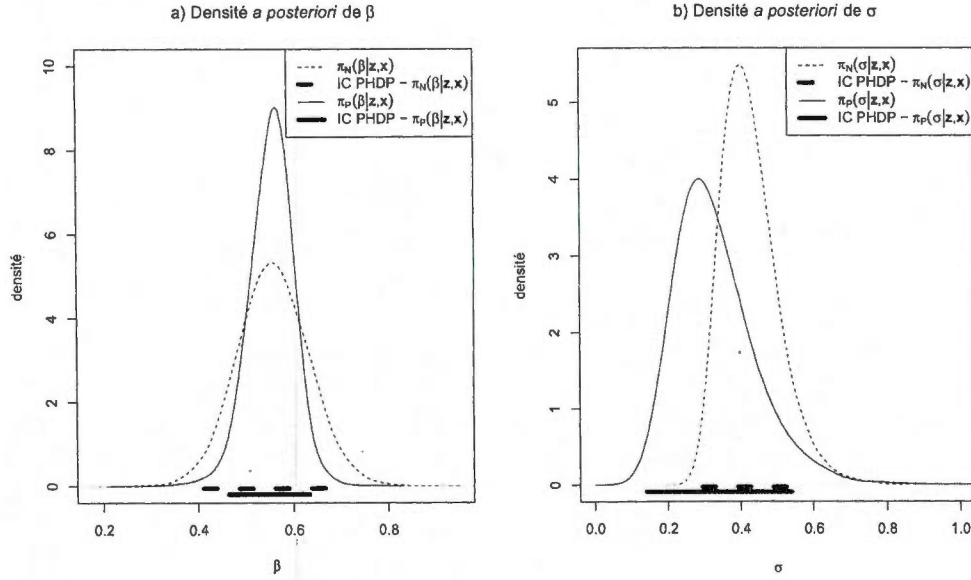


Figure 5.2 Illustration des densités *a posteriori* résultante.

la même que celle d'une loi normale centrée réduite, on s'attend uniquement à ce que les estimations de σ soient dans un même ordre de grandeur. Ils ne peuvent cependant pas être comparés.

En plus de l'estimation par la médiane et l'IC PHDP, un test d'hypothèse par la cote de Bayes est effectué. Il est de la forme $H_0 : \beta \neq 0$ versus $H_1 : \beta = 0$. Il permet de tester s'il y a un lien linéaire significatif ou non entre les deux variables étudiées.

Définition 10. La cote de Bayes permettant de déterminer l'évidence en faveur de $H_0 : \beta \neq 0$ est donnée par

$$\mathcal{B} = \frac{\Pr(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x} | H_0)}{\Pr(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x} | H_1)}.$$

On choisira H_0 si $\mathcal{B} \gg 1$. Dans notre contexte, \mathcal{B} est calculée de la manière suivante :

$$\begin{aligned}\mathcal{B} &= \frac{\Pr(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x} \mid \beta \neq 0)}{\Pr(\mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x} \mid \beta = 0)} \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) \Pr(\mathbf{X} = \mathbf{x})}{\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}, \beta = 0) \Pr(\mathbf{X} = \mathbf{x} \mid \beta = 0)} \\ &= \frac{\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x})}{\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}, \beta = 0)} \quad (\text{car } \beta \text{ est indépendant de } \mathbf{X}).\end{aligned}$$

Puisque X_i et Y_i sont continues, on utilise les densités. Alors,

$$\mathcal{B} = \frac{\int_{\beta} \int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma} f\left(\frac{y_i - \beta x_i}{\sigma}\right) d\sigma d\beta}{\int_{\sigma} \frac{1}{\sigma} \prod_{i=1}^n \frac{1}{\sigma} f\left(\frac{y_i}{\sigma}\right) d\sigma}.$$

Une échelle permettant d'utiliser la cote de Bayes est présentée au tableau 5.2. Elle a été établie à titre de référence par Jeffreys (1961).

\mathcal{B}	Preuve en faveur de H_0
$[0, 1)$	Négative (supporte H_1)
$[1, 3)$	Faible
$[3, 12)$	Positive
$[12, 150)$	Forte
$[150, \infty)$	Très forte

Tableau 5.2 Échelle de décision concernant la cote de Bayes.

Pour juillet 2010, la cote de Bayes est supérieure à 6.5×10^5 , à la fois pour l'hypothèse de normalité et celle d'une distribution DL-Pareto. Il y a donc un lien linéaire très fort entre les rendements quotidiens du S&P 500 et ceux du S&P/TSX pour juillet 2010. Cette conclusion est appuyée par le fait que les intervalles de crédibilité pour β ne contiennent pas 0.

L'analyse est maintenant effectuée sur un mois moins représentatif.

Les figures 5.3 et 5.4, où l'indice de la densité π indique la nature de la distribution des erreurs (normale ou DL-Pareto), suggèrent un lien linéaire positif entre les deux

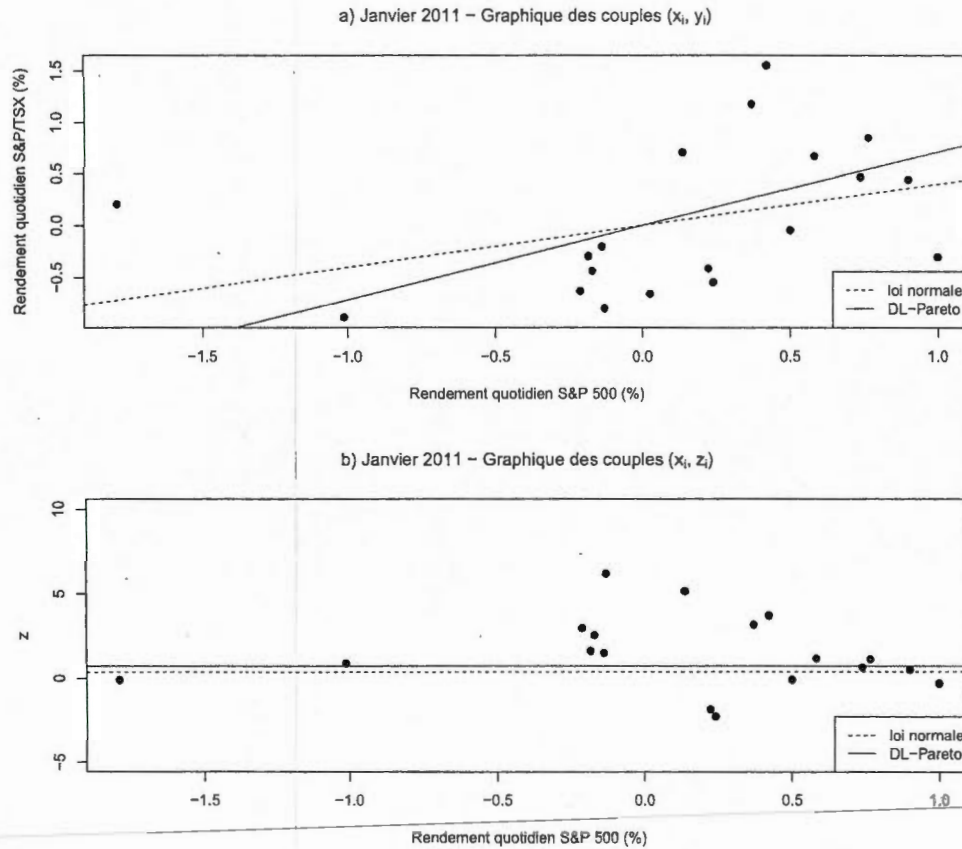


Figure 5.3 Illustration du modèle de régression où les droites sont estimées par la médiane *a posteriori*.

variables. Une hypothèse de distribution DL-Pareto sur les erreurs semble mieux traduire le comportement de la majorité des données. Dans ce cas, l'estimation de β par la médiane est égale à 0.71 versus 0.40 sous l'hypothèse de normalité. Toutefois, la largeur de l'intervalle de crédibilité PHDP de β est influencée par la variabilité des observations. Il est donné par $(0, 1.23)$ pour la distribution DL-Pareto et $(-0.1, 0.89)$ pour la loi normale. Pour ce qui est de σ , l'estimation par la médiane et l'IC PHDP sont respectivement égaux à 0.62 et $(0.29, 1.09)$ pour la distribution DL-Pareto, et 0.69 et $(0.48, 0.96)$ pour la loi normale. Le résultat du test $H_0 : \beta \neq 0$ versus $H_1 : \beta = 0$, pour janvier 2011, est donné par une cote de Bayes de 7 lorsque les erreurs sont distri-

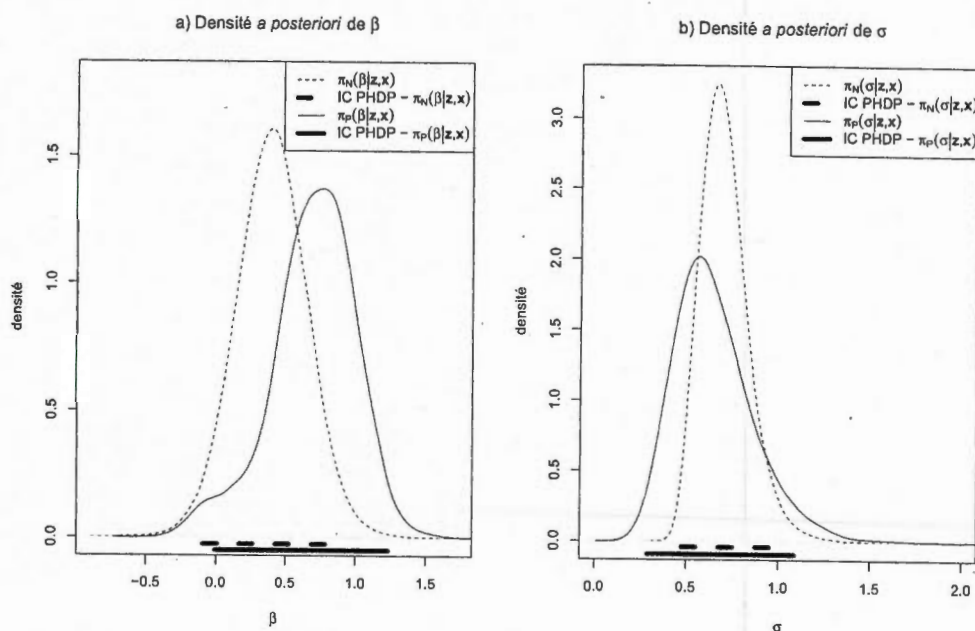


Figure 5.4 Illustration des densités *a posteriori* résultante.

buées selon une DL-Pareto puis 2.6 sous l'hypothèse de normalité. Ces résultats sont beaucoup moins favorables à H_0 que ceux obtenus pour juillet 2010. L'analyse faite en supposant la normalité des erreurs permettrait de conclure qu'il y a un lien linéaire faible entre les rendements quotidiens du S&P 500 et S&P/TSX pour janvier 2011 puisque la cote de Bayes est de 2.6. De plus, l'intervalle de crédibilité inclut 0. L'analyse faite en supposant une distribution DL-Pareto permet de conclure autrement. En effet, la cote de Bayes de 7 indique qu'il y a un lien linéaire positif entre les deux variables. De plus, la borne inférieure de l'intervalle de crédibilité à 95% pour β est 0, ce qui indique que cette valeur n'est pas très plausible.

Cette étude de cas a permis d'illustrer l'impact de la distribution du terme d'erreur. En effet, l'hypothèse de normalité peut mener à des décisions qui sont incohérentes avec la tendance décrite par les observations, lorsqu'on est en présence de valeurs aberrantes. En revanche, une hypothèse de distribution ayant des ailes suffisamment relevées engendre des résultats cohérents avec cette tendance général, même en présence de valeurs

aberrantes.

CONCLUSION

Dans ce mémoire, l'impact de la distribution du terme d'erreur sur l'inférence *a posteriori* a été étudié dans un cadre de régression linéaire simple. Il a été montré théoriquement que, sous une certaine condition relative aux ailes de la densité, l'inférence *a posteriori* est robuste face à des valeurs aberrantes en nombre minoritaire. Le modèle de régression qui a été utilisé peut être perçu comme un modèle de position-échelle. Vu de cet angle, un lien a pu être établi avec la théorie de robustesse de Desgagné (2011). Ensuite, la famille de densités, appelée DL-GEP, a été présentée afin d'offrir à l'utilisateur, par l'entremise de différents choix de paramètres, une panoplie de comportement d'ailes, dont ceux respectant le critère de robustesse. La distribution DL-Pareto en est un exemple. Elle a servi à illustrer les résultats théoriques à travers une analyse de données simulées et une étude de cas.

Bibliographie

- Barnes, C. 2010. « Bayesian nonparametric regression with a flexible error term distribution ». Mémoire de maîtrise, The University of Texas at El Paso.
- Box, G. E. P., et G. C. Tiao. 1968. « A bayesian approach to some outlier problems ». *Biometrika*, vol. 55, no. 1, p. 119–129. <<http://www.jstor.org/stable/2334456>>.
- Desgagné, A. 2011. Full robustness to outliers in a bayesian location-scale model. Rapport, Montréal, Québec, Canada, UQAM.
- Desgagné, A., et J.-F. Angers. 2007. « Conflicting information and location parameter inference ». *Metron*, vol. 65, no. 1, p. 67–97.
- Faul, A. C., et M. E. Tipping. 2001. « A variational approach to robust regression ». *Comp. Sci.*, vol. 2130, no. 2001, p. 95–102.
- Fonseca, T. C. O., M. A. R. Ferreira et H. S. Migon. 2008. « Objective bayesian analysis for the student-t regression model ». *Biometrika*, vol. 95, no. 2, p. 325–333. <<http://biomet.oxfordjournals.org/content/95/2/325.abstract>>.
- Jeffreys, H. 1961. *Theory of probability*. Jeffreys. Clarendon. <<http://books.google.ca/books?id=ZFnlSAAACAAJ>>.
- Kuss, M. 2006. « Gaussian process models for robust regression, classification, and reinforcement learning ». Thèse de Doctorat, Technische Universität Darmstadt.
- Liu, C. 1996. « Bayesian robust multivariate linear regression with incomplete data ». *J. Amer. Statist. Assoc.*, vol. 91, no. 435, p. 1219–1227.
- Mitra, K., A. Veeraraghavan et R. Chellappa. 2010. « Robust rvm regression using sparse outlier model ». In *2010 IEEE Conference on computer Vision and Pattern Recognition (CVPR)*, p. 1887–1894.

- Peña, D., R. Zamar et G. Yan. 2009. « Bayesian likelihood robustness in linear models ». *J. Stat. Plan. Infer.*, vol. 139, no. 7, p. 2196 – 2207. <<http://www.sciencedirect.com/science/article/pii/S0378375808004035>>.
- R Development Core Team. 2011. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Szpiro, A. A., K. M. Rice et T. Lumle. 2010. « Model-robust regression and a bayesian sandwich estimator ». *Ann. Appl. Stat.*, vol. 4, no. 4, p. 2099–2113.
- West, M. 1984. « Outlier models and prior distributions in bayesian linear regression ». *J. Roy. Stat. Soc. B Met.*, vol. 46, no. 3, p. 431–439.
- Yang, B., Z. Zhang et Z. Sun. 2007. « Robust relevance vector regression with trimmed likelihood function ». *Signal Processing Letters, IEEE*, vol. 14, no. 10, p. 746 –749.
- Yu, K., et R. A. Moyeed. 2001. « Bayesian quantile regression ». *Stat. Probabil. Lett.*, vol. 54, no. 4, p. 437–447.
- Zhu, H., J. G. Ibrahim et N. Tang. 2011. « Bayesian influence analysis : a geometric approach ». *Biometrika*, vol. 98, no. 2, p. 307–323. <<http://biomet.oxfordjournals.org/content/98/2/307.abstract>>.